

Kam, Chih-Min (AU1653)

To: Schreiber, David
Cc: Kam, Chih-Min (AU1653)
Subject: sequence search for 09/913,351

David:

Please do amino acid search with commercial data file. Thanks

Chih-Min Kam

NAME: Chih-Min Kam
AU: 1653
Date: 9/8/02
Room: 10D16
Mail Box: CM1, 9B01
Tel: 308-9437

Application No: 09/640,957.

Please search the following amino acid sequence(s) with commercial data file.

The claims cite a dimer (a conjugate) which has a minor subunit and a major subunit, where the minor subunit is

1. SEQ ID NO:2
2. SEQ ID NO:4

and the major subunit is

3. SEQ ID NO:3
4. SEQ ID NO:5

What is the best way to search this?

SEQ ID NO: 1 variable aa

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:39:28 : Search time 94.6 seconds
(without alignments)
43.443 Million cell updates/sec

Title: US-09-913-351-2
Perfect score: 186
Sequence: 1 ESKGREGSSQCRQEVQRKDLSCERTLRQSSRR 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186	100.0	37	21	AA15420
2	61.5	33.1	110	15	AA62762
3	61.5	33.1	140	10	AA62762
4	58	31.2	148	21	AA58050
5	58	31.2	148	21	AA58050
6	57.5	30.9	111	19	AA62664
7	57.5	30.9	111	19	AA62664
8	57	30.6	904	22	AA62664
9	55	29.6	156	20	AA40973
10	55	29.6	157	18	AA24164
11	55	29.6	157	18	AA24153

12	55	29.6	157	20	AA15245	Peanut allergen, A
13	55	29.6	157	22	AA04707	Anaphylactic antly
14	55	29.6	166	20	AA40968	Recombinant Ara h
15	55	29.6	207	22	AA82383	Peanut allergen Ar
16	54.5	29.3	165	22	ABG20170	Murine A20 Binding
17	54	29.0	430	21	AA44281	Castor bean 2S alb
18	53.5	28.8	34	21	AA15421	Drosophila melanog
19	53	28.5	593	22	AB58343	A Bcl-2 associated
20	53	28.5	917	21	AA53920	Novel human diagno
21	53	28.5	930	22	ABG16236	Novel human diagno
22	53	28.5	1450	22	ABG04173	Novel human diagno
23	53	28.5	1450	22	ABG07330	Novel human diagno
24	53	28.5	1462	22	AA45582	Human protein sequ
25	53	28.5	1475	22	AB11449	Human p73-kinase h
26	52.5	28.2	158	18	AA23586	Mablinin MBL1 from
27	52.5	28.2	158	18	AA23588	Mablinin MBL1 fr
28	52.5	28.2	758	22	ABG05454	Novel human diagno
29	52.5	28.2	758	22	ABG07092	Novel human diagno
30	52.5	28.2	758	22	ABG10119	Novel human diagno
31	52.5	28.2	758	22	ABG10239	Novel human diagno
32	52.5	28.2	758	22	ABG10412	Novel human diagno
33	52.5	28.2	758	22	ABG14939	Novel human diagno
34	52.5	28.2	797	22	ABG19906	Novel human diagno
35	52	28.0	1775	22	AB564008	Drosophila melanog
36	52	28.0	2061	22	AB566938	Drosophila melanog
37	51.5	27.7	91	22	AA673889	Human colon cancer
38	51.5	27.7	157	19	AA53261	Amino acid sequenc
39	51.5	27.7	171	15	AA53579	Synthetic 2S seed
40	51.5	27.7	174	22	AA72901	Flax 2S storage pr
41	51.5	27.4	52	22	AB52953	Peptide #2204 enco
42	51	27.4	52	22	AB32337	Peptide #4988 enco
43	51	27.4	52	22	AB34736	Peptide #2242 enco
44	51	27.4	52	22	AB37585	Peptide #5101 enco
45	51	27.4	52	22	AB22889	Protein #488 enco

ALIGNMENTS

RESULT	1	
AA15420	AA15420 standard; peptide; 37 AA.	
XX		
XX	AA15420;	
AC		
XX		
DT	17-JAN-2001 (first entry)	
XX		
DE	Castor bean 2S albumin storage protein amino acids 36-72.	
XX		
KW	Antitumour; antimicrobial; immunostimulatory; glycoconjugate; mannose;	
KW	polysaccharide; mannose; galactose; castor bean; immune response; human;	
KW	2S albumin storage protein; animal; tumour necrosis factor; mononuclear;	
KW	phagocyte; granulocyte; macrophage; tumour; immunogen; Candida utilis.	
XX		
OS	Ricinus communis.	
XX		
PN	WO200050087-A1.	
XX		
PD	31-AUG-2000.	
XX		
PF	21-OCT-1999; 99WO-ES00338.	
XX		
PR	26-FEB-1999; 99ES-0000408.	
XX		
PA	(INFA-) IND FARM CANTABRIA SA.	
XX		
PI	Brieva Delgado A, Garcia Villarrubia V, Guerrero Gomez-Pamo A;	
PI	Pivel Ranieri JP, Gimenez Gallego G, Matji Tuduri JA;	
XX		
DR	WPI; 2000-558369/51.	
XX		
PT	New glycoconjugate, useful for treating immunological disorders,	
PT	comprises polysaccharide from candida utilis and polypeptide from	

CC a seed-specific promoter and expressed at high levels only or mostly
CC in the seed forming stage and produced mostly in the seeds.
XX
SQ Sequence 140 AA;

Query Match 33.1%; Score 61.5; DB 10; Length 140;
Best Local Similarity 40.7%; Pred. No. 1.2;
Matches 11; Conservative 8; Mismatches 7; Indels 1; Gaps 1;

OY 7 EGSSSQCCQEQVQRKD-LSCEERYLRQ 32
I : :|||::||: || | :||
Db 27 eeengeecreqmrgqmlshcrmyrmq 53

RESULT 4
AAV58049
ID AAV58049 standard; Protein; 148 AA.

AC AAV58049;

DT 14-MAR-2000 (first entry)

DE Puroindoline B protein sequence.

KW Puroindoline B; friabillin; grain softness; wheat; starch; grain hardness;
KW grain texture; transgenic plant; sorghum; rice; barley; maize; rye;
KM oat; triticale.

OS Triticum aestivum.

PN WO9961580-A2.

PD 02-DEC-1999.

PF 21-MAY-1999; 99WO-US11379.

PR 22-MAY-1998; 98US-0083852.

PI (RERE-) RES & DEV INST INC.

PA (USDA) US DEPT OF AGRICULTURE.

PI Morris CF, Giroux MJ;

DR WPI: 2000-062702/05.

DR N-PSDB: AAZ47520.

PT Altering grain hardness in plants by modifying expression of
PT puroindoline proteins -
XX Examples; Page 62; 75pp; English.

CC This is the wheat puroindoline B protein sequence. Friabillin is a 15kD
CC marker protein for grain softness, and is composed of Puroindoline A and
CC Puroindoline B. Friabillin is present in high amounts on the surface of
CC soft wheat starch, and in low amounts on hard wheat starch. The
CC puroindolines have tryptophan-rich, hydrophobic domains which have
CC affinity for binding lipids. Puroindoline A and puroindoline B are used
CC in the production of transgenic plants with modified grain hardness. The
CC invention can be used to produce plants with softer grain texture, where
CC at least one parent plant has a hard texture grain. The softer grained
CC plants are produced through the introduction of a nucleic acid sequence
CC which operably encodes a puroindoline protein into a cell from the parent
CC plant, progeny plants are generated for the cell containing the nucleic
CC acid sequence. The methods can be used for producing transgenic plants
CC with softer textured grain from plants such as durum wheat, sorghum,
CC rice, barley and maize. They can also be used for producing transgenic
CC plants with harder textured grain from plants such as wheat, rye,
CC triticale and oat. The modification of the grain texture results in
CC softer grains which require less milling.

SQ Sequence 148 AA;

Query Match 31.2%; Score 58; DB 21; Length 148;
Best Local Similarity 52.0%; Pred. No. 3.8;
Matches 13; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

OY 8 GSSSQCCQEQVQRKD-LSCEERYLRQ 32
I ||||| || | ||||| : :
Db 34 g9sgsqcqqe--rpkissckdyyme 56

RESULT 5
AAV58050
ID AAV58050 standard; Protein; 148 AA.

AC AAV58050;

DT 14-MAR-2000 (first entry)

DE Serine substituted puroindoline B protein sequence.

KW Puroindoline B; friabillin; grain softness; wheat; starch; grain hardness;
KW grain texture; transgenic plant; sorghum; rice; barley; maize; rye;
KM oat; triticale.

OS Triticum aestivum.

PN WO9961580-A2.

PD 02-DEC-1999.

PF 21-MAY-1999; 99WO-US11379.

PR 22-MAY-1998; 98US-0083852.

PI (RERE-) RES & DEV INST INC.

PA (USDA) US DEPT OF AGRICULTURE.

PI Morris CF, Giroux MJ;

DR WPI: 2000-062702/05.

DR N-PSDB: AAZ47521.

PT Altering grain hardness in plants by modifying expression of
PT puroindoline proteins -
XX Disclosure; Page 63; 75pp; English.

CC This is the wheat serine substituted puroindoline B protein sequence.
CC Friabillin is a 15kD marker protein for grain softness, and is composed of
CC Puroindoline A and puroindoline B. Friabillin is present in high amounts
CC on the surface of soft wheat starch, and in low amounts on hard wheat
CC starch. The puroindolines have tryptophan-rich, hydrophobic domains which
CC have affinity for binding lipids. Puroindoline A and puroindoline B are
CC used in the production of transgenic plants with modified grain hardness.
CC The invention can be used to produce plants with softer grain texture,
CC where at least one parent plant has a hard texture grain. The softer
CC grained plants are produced through the introduction of a nucleic acid
CC sequence which operably encodes a puroindoline protein into a cell from
CC the parent plant, progeny plants are generated for the cell containing
CC the nucleic acid sequence. The methods can be used for producing
CC transgenic plants with softer textured grain from plants such as durum
CC wheat, sorghum, rice, barley and maize. They can also be used for
CC producing transgenic plants with harder textured grain from plants such
CC as wheat, rye, triticale and oat. The modification of the grain texture
CC results in softer grains which require less milling.

SQ Sequence 148 AA;

Query Match 31.2%; Score 58; DB 21; Length 148;
Best Local Similarity 52.0%; Pred. No. 3.8;
Matches 13; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

PN	MO9724139-A1.
XX	
PD	10-JUL-1997.
XX	
PE	23-SEP-1996; 96WO-US15222.
XX	
PR	04-MAR-1996; 96US-0610424.
PR	29-DEC-1995; 95US-0009455.
XX	
PA	(UYAR-) UNIV ARKANSAS.
PI	Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;
XX	
XX	WPI; 1997-363453/33.
DR	N-PSDB; AAT76614.
XX	
PT	Peanut allergens Ara hi and Ara hII - used for vaccination and in
PT	two-site monoclonal antibody based ELISA
XX	
PS	Claim 31; Page 219; 354pp; English.
XX	
CC	This polypeptide comprises major peanut allergen Ara hII.
CC	Its sequence was deduced from cDNA clone p38 (AAT76614), isolated
CC	from peanut seed cDNA using a primer (see AAT76617) based on an
CC	isolated Ara hi peptide (see AAW24151). The sequence shows
CC	significant homology with the conglutinin family of seed storage
CC	proteins of other legumes. The allergen is recognised by serum
CC	IgE from a large proportion of individuals with peanut
CC	hypersensitivity. Ara hII and Ara hi (see AAW24149-50) can be used to
CC	raise monoclonal antibodies which are used in a specific two-site
CC	MAb ELISA for the detection of Ara hi or Ara hII (claimed). IgE-
CC	binding Ara hII antigen epitopes (see AAW24188-93) may be used in
CC	vaccines to protect against allergic reactions to peanut allergens,
CC	e.g. anaphylactic shock.
XX	
XX	Sequence 157 AA;

```

Query Match 29.6%; Score 55; DB 18; Length 157;
Best Local Similarity 30.6%; Pred. NO. 10;
Matches 11; Conservative 10; Mismatches 9; Indels 6; Gaps 1.

OY      1  ESKGEREGSSQOCROEVQRKDLSSCERYLRSSSR 36
      1  : : : : : : : : : : : : : : : :
Db      23  elggr-----rcqsgleranlrpceqhlmqkigr 52

RESULT 11
AAW24153
ID      AAW24153 standard; Protein: 157 AA.
XX
XX      AAW24153;
AC
XX
XX      29-DEC-1997 (first entry)
DT
XX
XX      Peanut allergen Ara hII.
DE
XX
XX      Peanut; seed storage protein; allergen; allergy; hypersensitivity;
KW      vaccine; anaphylactic shock; immunotherapy; therapy;
KW      monoclonal antibody; ELISA; analysis; Ara hII.
XX
XX      Arachis hypogaea strain Florunner.
OS
XX
XX      WO9724139-A1.
PN
XX      10-JUL-1997.
PD
XX
XX      23-SEP-1996; 96WO-US15222.
PE
XX
XX      04-MAR-1996; 96US-0610424.
PR      29-DEC-1993; 95US-0009455.
XX
XX      (UYAR-) UNIV ARKANSAS.
XX

```

XX	Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS:
PI	
XX	WPI: 1997-363453/33.
DR	N-PSDB: AAT76615.
XX	
PT	peanut allergens Ara hi and Ara hII - used for vaccination and in
XX	two-site monoclonal antibody based ELISA
PS	
XX	Claim 31; Page 198; 354pp; English.
XX	
CC	This polypeptide comprises major peanut allergen Ara hII.
CC	Its sequence was deduced from a cDNA clone (AAT76615) isolated
CC	from peanut seed cDNA using a primer (see AAT76617) based on an
CC	isolated Ara hi peptide (see AAW24151). The sequence shows
CC	significant homology with the conglutin family of seed storage
CC	proteins of other legumes. The allergen is recognised by serum
CC	IgE from a large proportion of individuals with peanut
CC	hypersensitivity. Ara hII and Ara hi (see AAW24149-50) can be used to
CC	raise monoclonal antibodies which are used in a specific two-site
CC	MAb ELISA for the detection of Ara hi or Ara hII (claimed). IgE-
CC	binding Ara hII antigen epitopes (see AAW24188-93) may be used in
CC	vaccines to protect against allergic reactions to peanut allergens,
CC	e.g. anaphylactic shock.
XX	
XX	Sequence 157 AA;
XX	

```

Query Match      29.6%; Score 55; Db 18; Length 157;
Best Local Similarity 30.6%; Pred. No. 10;
Matches 11; Conservative 10; Mismatches 9; Indels 6; Gaps 1;

QY      1  ESKGEREGSSQCGREQVEYVRKDLSCERYLRQSSSR 36
      1  : : : : : : : : : : : : : : : : : :
Db      23  elgqdr-----rcgsqteranlrpceqhlmgkixqr 52

RESULT 12
AAV15245
ID      AAV15245 standard; Protein: 157 AA.
XX      AAV15245;
AC      AAV15245;
XX      AAV15245;
DT      09-NOV-1999 (first entry)
DE      Peanut allergen, Ara h 2, amino acid sequence.
XX      allergen; Immune response; transgenic; allergen; epitope;
KW      immunoglobulin E; Ig E; binding site; peanut.
XX      Arachis hypogea.
OS      Arachis hypogea.
XX      WO9398978-A1.
XX      WO9398978-A1.
XX      05-AUG-1999.
XX      05-AUG-1999.
XX      29-JAN-1999; 99WO-US02031.
XX      27-AUG-1998; 98US-0141220.
XX      31-JAN-1998; 98US-0073283.
XX      13-FEB-1998; 98US-0074590.
XX      13-FEB-1998; 98US-0074624.
XX      13-FEB-1998; 98US-0074633.
XX      (SOSI/) SOSIN H.
XX      (SOSI-) UNIT ARKANSAS.
XX      (UYNY ) UNIV NEW YORK MT SINAI SCHOOL MEDICINE.
XX      Bannon GA, Burks AW, Sampson HA, Sosin H;
XX      WPI; 1999-479188/40.
XX      N-PSDB; AAZ06383.
XX

```


Query Match 29.6%; Score 55; DB 20; Length 166;
 Best Local Similarity 30.6%; Pred. No. 11;
 Matches 11; Conservative 10; Mismatches 9; Indels 6; Gaps 1;
 Oy 1 ESKGEREGSSSQOCROEVQKRLSCERYLRQSSSR 36
 Db 21 elqgdr-----rcsqqlerantlrpceqhlmqkigr 50

Search completed: September 9, 2002, 12:42:39
 Job time: 191 sec

RESULT 15
 AAB82383
 ID AAB82383 standard; Protein; 207 AA.
 XX
 AC AAB82383;
 XX
 DT 23-JUL-2001 (first entry)
 XX
 DE Peanut allergen Ara h2 gene product.
 XX
 KW Peanut; allergen; Ara h2; transgenic plant; allergy.
 XX
 OS Arachis hypogea.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= signal_peptide
 FT 22..207
 FT Protein /label= Mature_protein
 XX
 PN W0200136621-A2.
 XX
 PD 25-MAY-2001.
 XX
 PE 20-NOV-2000; 2000WO-US31657.
 XX
 PR 19-NOV-1999; 99US-0167255.
 XX
 PA (UYAL-) UNTV ALABAMA A & M.
 XX
 PI Dodo HW, Arntzen CJ, Konan KN, Viquez OM;
 XX
 DR WPI; 2001-355630/37.
 DR N-PSDB; AAF90336, AAF90337.
 XX
 PT Producing transgenic peanut plants that produce allergen-free seeds,
 PT useful in non-allergenic foods, by antisense or sense co-suppression of
 PT allergen-encoding genes -
 XX
 PS Example 1; Fig 2; 72pp; English.
 XX
 CC The present sequence is that of the peanut allergenic protein (AP)
 CC encoded by the Ara h2 gene (see AAF90336). The invention relates
 CC to a method for producing a peanut plant having reduced, or
 CC undetectable, AP content in its seed. A peanut plant cell is
 CC transformed with a DNA construct containing an antisense AP gene
 CC and/or sense AP gene, or their fragments, regenerated to plants,
 CC and fertile transgenic plants that produce seeds with reduced AP
 CC content are identified. The AP sense or antisense gene may
 CC comprise at least a portion of the Ara h2 gene sequence. The seeds
 CC are useful for preparation of allergen-free foods. Recombinant AP
 CC may be produced and used to produce antibodies useful for detecting
 CC AP in foods, and for treatment or prevention of peanut allergy.
 CC
 XX
 SQ Sequence 207 AA;

Query Match 29.6%; Score 55; DB 22; Length 207;
 Best Local Similarity 30.6%; Pred. No. 14;
 Matches 11; Conservative 10; Mismatches 9; Indels 6; Gaps 1;

Oy 1 ESKGEREGSSSQOCROEVQKRLSCERYLRQSSSR 36
 Db 26 elqgdr-----rcsqqlerantlrpceqhlmqkigr 55

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:40:18 : Search time 34.92 seconds
(without alignments)
25.880 Million cell updates/sec

Title: US-09-913-351-2

Perfect score: 186
Sequence: 1 ESKGERGSSSQCRQEVORRDLSCERYLRSSSR 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5C_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5D_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5E_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/5F_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52.5	28.2	158	2	US-08-670-186-4
2	52.5	28.2	158	2	US-08-670-186-6
3	51.5	27.7	157	2	US-08-706-391B-7
4	50	26.9	404	3	US-08-911-853-9
5	50	26.9	404	4	US-09-479-409-9
6	50	26.9	404	4	US-09-479-453-9
7	49	26.3	873	3	US-08-990-140-2
8	49	26.3	873	3	US-09-546-238-2
9	48.5	26.1	314	1	US-08-444-231-19
10	48.5	26.1	314	1	US-08-152-443A-19
11	48.5	26.1	314	5	PCT-US95-17083-4
12	48	25.8	486	2	US-08-942-423-2
13	48	25.8	486	4	US-08-630-915A-26
14	48	25.8	1898	1	US-08-056-200-94
15	48	25.8	1898	1	US-08-800-644-94
16	47.5	25.5	158	2	US-08-618-911-4
17	47.5	25.5	158	3	US-08-938-675A-2
18	47.5	25.5	235	1	US-08-928-443-4
19	47.5	25.5	235	1	US-09-129-055-4
20	47	25.3	174	4	US-09-383-586-12
21	47	25.3	186	2	US-08-557-309B-43
22	47	25.3	186	3	US-08-834-306-43
23	46.5	25.0	186	4	US-08-993-674A-43
24	46.5	25.0	514	4	US-08-796-899-25
25	46.5	24.7	308	2	US-08-807-050-4
26	45.5	24.5	360	2	US-08-531-927B-2
27	45.5	24.5	360	4	US-09-041-886-13

28	45	24.2	161	1	US-08-948-195-3	Sequence 3, Appl1
29	45	24.2	161	2	US-09-094-080-3	Sequence 3, Appl1
30	45	24.2	161	2	US-09-216-625-3	Sequence 3, Appl1
31	45	24.2	193	2	US-08-383-621-3	Sequence 3, Appl1
32	45	24.2	193	3	US-08-459-906-3	Sequence 2, Appl1
33	45	24.2	332	4	US-09-026-408-2	Sequence 2, Appl1
34	45	24.2	405	4	US-09-026-408-13	Sequence 13, Appl1
35	45	24.2	600	1	US-08-253-785-3	Sequence 3, Appl1
36	44	23.7	27	4	US-09-245-712-12	Sequence 12, Appl1
37	44	23.7	27	4	US-09-245-712-13	Sequence 13, Appl1
38	44	23.7	27	4	US-09-245-712-14	Sequence 14, Appl1
39	44	23.7	191	1	US-07-885-689A-29	Sequence 29, Appl1
40	44	23.7	191	1	US-08-093-383-3	Sequence 3, Appl1
41	44	23.7	193	2	US-08-383-621-2	Sequence 2, Appl1
42	44	23.7	193	3	US-08-459-906-2	Sequence 2, Appl1
43	44	23.7	199	1	US-07-801-164A-4	Sequence 4, Appl1
44	44	23.7	217	2	US-09-105-651-2	Sequence 2, Appl1
45	44	23.7	437	1	US-08-810-116-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1
US-08-670-186-4
Sequence 4, Application US/08670186
Patent No. 5859343
GENERAL INFORMATION:
APPLICANT: SUN, SAMUEL S.M.
APPLICANT: XIONG, LIWEN
APPLICANT: HU, ZHONG
APPLICANT: CHEN, HANG
TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MARILIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,186
FILING DATE: 21-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MORASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 23461-20007.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSFOERSWSH
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-670-186-4

Query Match 28.2% Score 52.5; DB 2; Length 158;
Best Local Similarity 27.3%; Pred. No. 5.1;
Matches 9; Conservative 12; Mismatches 1; Gaps 1;
OY 5 EREGSSSQCRQEVORRDLSCERYLRSSSR 36
: : : |::|:|:|:|:|:|:|

Db 33 DNDDENPICRFOFOOHHRACQYIRRAQR 65

RESULT 2
US-08-670-186-6
; Sequence 6, Application US/08670186
; Patent No. 5859343
; GENERAL INFORMATION:
; APPLICANT: SUN, SAMUEL S.M.
; APPLICANT: XIONG, LIWEN
; APPLICANT: HU, ZHONG
; APPLICANT: CHEN, HANG
; TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,186
; FILING DATE: 21-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 23461-20007.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 822-0168
; TELEEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-670-186-6

Query Match 28.2%; Score 52.5; DB 2; Length 158;
Best Local Similarity 27.3%; Pred. No. 5.1;
Matches 9; Conservative 12; Mismatches 11; Indels 1; Gaps 1;

OY 5 EREGSSSQOCROEVOR-KDLSSCERLRSSSR 36
Db 33 DNDDENPICRFOFOOHHRACQYIRRAQR 65

RESULT 3
US-08-706-391B-7
; Sequence 7, Application US/08706391B
; Patent No. 6174725
; GENERAL INFORMATION:
; APPLICANT: ANDERSON, OLIN D
; TITLE OF INVENTION: ALTERING DOUGH VISCOELASTICITY WITH
; MODIFIED GLUTENINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: USDA, AGRICULTURAL RESEARCH SERVICE, PACIFIC
; WEST
; STREET: 800 BUCHANAN STREET
; CITY: ALBANY
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94710

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,391B
; FILING DATE: 30-Aug-1996
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNOR, MARGARET A
; REGISTRATION NUMBER: 30,043
; REFERENCE/DOCKET NUMBER: 0235.95/USDA96-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 559-6067
; TELEFAX: (510) 559-5777
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 157 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-706-391B-7

Query Match 27.7%; Score 51.5; DB 4; Length 157;
Best Local Similarity 35.5%; Pred. No. 6.9;
Matches 11; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

OY 5 EREGSSSQOC--ROEVORRDLSSCERYLRQ 32
Db 2 EGEASELOCDRELQELQERELKACQOVMDQ 32

RESULT 4
US-08-911-853-9
; Sequence 9, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Geritise, Gijtsbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,853
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:

LENGTH: 404 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-911-853-9

Query Match 26.9%; Score 50; DB 3; Length 404;
Best Local Similarity 45.8%; Pred. No. 32;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 ESKGEREGSSQOCROEVORKDLS 24
DB 15 QOKGVLEGSARQVRLDRKDLS 38

RESULT 5
US-09-479-409-9
Sequence 9, Application US/09479409
Patent No. 6225106

GENERAL INFORMATION:
APPLICANT: Gerritse, Gijtsbert
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/479,409
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 404 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-479-409-9

Query Match 26.9%; Score 50; DB 4; Length 404;
Best Local Similarity 45.8%; Pred. No. 32;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 ESKGEREGSSQOCROEVORKDLS 24
DB 15 QOKGVLEGSARQVRLDRKDLS 38

RESULT 6
US-09-479-453-9
Sequence 9, Application US/09479453
Patent No. 6313283

GENERAL INFORMATION:
APPLICANT: Gerritse, Gijtsbert
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/479,453
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 404 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-479-453-9

Query Match 26.9%; Score 50; DB 4; Length 404;
Best Local Similarity 45.8%; Pred. No. 32;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 ESKGEREGSSQOCROEVORKDLS 24
DB 15 QOKGVLEGSARQVRLDRKDLS 38

RESULT 7
US-08-990-140-2
Sequence 2, Application US/08990140A
Patent No. 6093795

GENERAL INFORMATION:
APPLICANT: Olsen, Henrik S.
APPLICANT: Ruben, Steven M.
APPLICANT: Wetli, Nathalie
APPLICANT: Rom, Eran
TITLE OF INVENTION: Human Prt1-like Subunit Protein (hprt1) and Human
TITLE OF INVENTION: elf4G-like Protein (p97) Genes
FILE REFERENCE: 1488, 0700001
CURRENT FILING DATE: 1997-12-12
EARLIER FILING DATE: 1996-12-13
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 873
TYPE: PRT
ORGANISM: Homo sapiens
US-08-990-140-2

Query Match 26.3%; Score 49; DB 3; Length 873;
Best Local Similarity 44.4%; Pred. No. 1e+02;
Matches 12; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 4 GREGSSSQCRQEVQRDLSCERYL 30
DB 362 GEKFKQIQRFHQGVOLIDFSCERYL 388

RESULT 8

US-09-546-238-2
Sequence 2, Application US/09546238
Patent No. 6316225
GENERAL INFORMATION:
APPLICANT: Olsen, Henrik S.
APPLICANT: Ruben, Steven M.
APPLICANT: Sonenberg, Nahum
APPLICANT: Method, Nathalie
APPLICANT: Rom, Eran
TITLE OF INVENTION: Human Pti1-like Subunit Protein (Pti1) Polynucleotides
TITLE OF INVENTION: (as amended)
FILE REFERENCE: 1488.0700002
CURRENT APPLICATION NUMBER: US/09/546,238
CURRENT FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: US 60/033,151
PRIOR FILING DATE: 1996-12-13
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 873
TYPE: PRT
ORGANISM: Homo sapiens
US-09-546-238-2

Query Match 26.3%; Score 49; DB 4; Length 873;
Best Local Similarity 44.4%; Pred. No. 1e+02;
Matches 12; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 4 GREGSSSQCRQEVQRDLSCERYL 30
DB 362 GEKFKQIQRFHQGVOLIDFSCERYL 388

RESULT 9

US-08-444-231-19
Sequence 19, Application US/08444231
Patent No. 5652210
GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J.
APPLICANT: SHAPIRO, JOHN P.
APPLICANT: KIEFER, MICHAEL C.
TITLE OF INVENTION: NOVEL FAS PROTEIN AND METHODS OF USE
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,231
FILING DATE: 18-MAY-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/152,443
FILING DATE: 15-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: LEHNHARDT, SUSAN K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20006.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-231-19

Query Match 26.1%; Score 48.5; DB 1; Length 314;
Best Local Similarity 26.9%; Pred. No. 39;
Matches 7; Conservative 14; Mismatches 4; Indels 1; Gaps 1;

QY 9 SSSQCCRQEVQRDL-SSCERYLRQS 33
DB 160 TSNTRCKREVKRKEVQKCRKRKEN 185

RESULT 10

US-08-152-443A-19
Sequence 19, Application US/08152443A
Patent No. 5663070
GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J.
APPLICANT: SHAPIRO, JOHN P.
APPLICANT: KIEFER, MICHAEL C.
TITLE OF INVENTION: NOVEL FAS PROTEIN AND METHODS OF USE
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/152,443A
FILING DATE: 15-NOV-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LEHNHARDT, SUSAN K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20006.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-152-443A-19

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: September 9, 2002, 12:40:58 ; Search time 22.97 Seconds
(without alignments)
62.369 Million cell updates/sec

Title: US-09-913-351-2

Perfect score: 186
Sequence: 1 ESKGEREGSSSQOCROEVORKDLSCCERYLROSSRR 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186	100.0	258	1 2SS_RICCO	P01089 ricinus com
2	62	33.3	749	1 SWAP_CAEEL	Q10580 caenorhabdi
3	61.5	33.1	146	1 2SS_BEREX	P04403 bertolletti
4	59	31.7	295	1 2SS5_HELAN	P15461 helianthus
5	58	31.2	148	1 PUTB_WHEAT	Q10464 triticum ae
6	56	30.1	37	1 CG2S_LUPAN	P09930 lupinus ang
7	55	29.6	648	1 GLUT_WHEAT	P10387 triticum ae
8	55	29.6	660	1 GLUT3_WHEAT	P08488 triticum ae
9	55	29.6	761	1 Y218_HUMAN	Q93075 homo sapien
10	53.5	28.8	100	1 2SS4_CAPMA	P80353 capparis ma
11	53.5	28.8	104	1 2SS3_CAPMA	P08489 triticum ma
12	53.5	28.8	838	1 GLUT_WHEAT	P10387 triticum ae
13	53	28.5	839	1 GLUT5_WHEAT	P10388 triticum ae
14	53	28.5	215	1 RNI_AERY	Q07465 aeromonas h
15	53	28.5	518	1 GSH1_BUCAL	P57485 buchiera ap
16	53	28.5	1448	1 PK3G_HUMAN	O75747 homo sapien
17	52	28.0	87	1 RL37_SCHMA	Q44125 schistosoma
18	51.5	27.7	104	1 2SS1_CAPMA	P80351 capparis ma
19	51	27.4	524	1 SBP_SOYBN	Q04672 glycine max
20	51	27.4	701	1 ORA_PLAFN	P04672 glycine max
21	50	26.9	557	1 YC98_SCHPO	Q09845 plasmodium
22	50	26.9	1704	1 VITI_FUNHE	Q09845 plasmodium
23	49.5	26.6	186	1 G119_ORYSA	Q05058 fundulus he
24	49.5	26.6	215	1 SC2_OCTDO	P29835 oryza sativ
25	49	26.3	307	1 COPE_HUMAN	P27010 octopus dof
26	49	26.3	489	1 DMP1_RAT	Q14579 homo sapien
27	49	26.3	494	1 VE2_HPV12	P81833 rattus norv
28	49	26.3	300	1 IF39_HUMAN	P26782 human papil
29	48.5	26.1	300	1 HAP2_KLILA	P53844 homo sapien
30	48.5	26.1	606	1 HMD2_DROAN	P32554 drosophila
31	48.5	26.1	999	1 CARE_MOUSE	Q09845 plasmodium
32	48	25.8	486	1 HSI_MOUSE	P49710 mus musculu
33	48	25.8	562	1 AMAL_PLAFR	P22622 plasmodium

34	48	25.8	815	1 GTRB_MYXXA	O33367 myxococcus
35	48	25.8	1004	1 CARE_HUMAN	Q09x16 homo sapien
36	48	25.8	1506	1 PK3G_MOUSE	O70167 mus musculu
37	48	25.8	1898	1 TRHY_HUMAN	O07283 homo sapien
38	48	25.8	4473	1 PIEL_CRIGR	O91155 cricetus
39	48	25.8	4687	1 PIEL_RAT	P30427 rattus norv
40	47.5	25.5	82	1 Y467_TREPA	O83480 treponema p
41	47.5	25.5	110	1 2SS1_BRANA	P24565 brassica na
42	47.5	25.5	158	1 2SS5_SOYBN	P19594 glycine max
43	47.5	25.5	235	1 ARDH_HUMAN	P41227 homo sapien
44	47.5	25.5	487	1 MDM2_CANPA	P56950 canis famli
45	47	25.3	376	1 DUT_HSV62	P52541 human herpe

ALIGNMENTS

RESULT 1
ID 2SS_RICCO STANDARD: PRT; 258 AA.
AC P01089;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S albumin precursor (Allergen Ric c 1).
OS Ricinus communis (Castor bean).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91105729; PubMed=2274038;
RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor: a single preprotein
may be processed into two different heterodimeric storage proteins.";
RL Mol. Gen. Genet. 222:400-408(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91016940; PubMed=2216785;
RA Irwin S.D., Lord J.M.;
RT "Nucleotide sequence of a Ricinus communis 2S albumin precursor
gene.";
RL Nucleic Acids Res. 18:5890-5890(1990).
RN [3]
RP SEQUENCE OF 157-190 AND 194-258.
RX MEDLINE=83082772; PubMed=7174664;
RA Sharief F.S., Li S.S.-L.;
RT "Amino acid sequence of small and large subunits of seed storage
protein from Ricinus communis.";
RL J. Biol. Chem. 257:14753-14759(1982).
RN [4]
RP SIMILARITY TO PROTEINASE INHIBITORS.
RX MEDLINE=83308577; PubMed=6615448;
RA Odani S., Koide T., Ono T., Ohnishi K.;
RT "Structural relationship between barley (Hordeum vulgare) trypsin
inhibitor and castor-bean (Ricinus communis) storage protein.";
RL Biochem. J. 213:543-545(1983).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
LINKED BY 2 DISULFIDE BONDS.
CC -1- PM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE
CHAINS INVOLVE CYS-162 AND CYS-175.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54158; CAA38097.1; -.
CC DR PIR; A01328; RECS.
CC DR PIR; S11499; S11499.
CC DR PIR; S11500; S11500.
CC DR PIR; S11501; S11501.
CC DR InterPro; IPR003612; AAI.
CC DR InterPro; IPR001768; Cereals_1ryp_aml_1nh.
CC DR Pfam; PF00234; try_1ryp_aml_1. 2.
CC DR SMART; SM00499; AAI; 2.
CC KM Seed storage protein; Signal; Allergen.
CC FT SIGNAL 1 21 PROBABLE.
CC FT PROPEP 22 156
CC FT CHAIN 157 190 2S ALBUMIN, SMALL CHAIN.
CC FT PROPEP 191 193
CC FT CHAIN 194 258
CC FT MOD_RES 194 194 2S ALBUMIN, LARGE CHAIN.
CC FT CONFLICT 222 222 PYROLIDONE CARBOXYLIC ACID.
CC FT CONFLICT 226 229 E -> Q (IN REF. 3).
CC FT CONFLICT 234 234 MISSING (IN REF. 3).
CC FT CONFLICT 255 255 D -> N (IN REF. 3).
CC FT CONFLICT 255 255 E -> Q (IN REF. 3).
CC SQ SEQUENCE 258 AA; 29290 MW; 27874CFC50E41072 CRC64;

Query Match 100.0%; Score 186; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 2.1e-16;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ESKGEREGSSQOCROEVORLSSCERYLRROSSRR 37
Db 36 ESKGEREGSSQOCROEVORLSSCERYLRROSSRR 72

RESULT 2
SWAP_CAEEL STANDARD; PRT; 749 AA.
AC Q10380;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SWAP protein (Suppressor of white apricot protein homolog).
GN SWAP.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBL_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95061415; PubMed=7971282;
RA Spikes D.A., Kramer J., Bingham P.M., van Doren K.;
RT "SWAP pre-mRNA splicing regulators are a novel, ancient protein
RT family sharing a highly conserved sequence motif with the prp1
RT family of constitutive splicing proteins.";
RL Nucleic Acids Res. 22:4510-4519(1994).
CC -1- FUNCTION: IT IS A REGULATOR OF PRE-MRNA SPLICING (AND, POSSIBLY,
CC OF OTHER RNA PROCESSING EVENTS). IT MAY REGULATE ITS OWN
CC EXPRESSION AT THE LEVEL OF RNA PROCESSING.
CC -1- SIMILARITY: TO DROSOPHILA SWA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U06932; AAA64937.1; -.
CC EMBL; U06933; AAA64938.1; -.
CC DR InterPro; IPR000061; Surp. 2.
CC DR Pfam; PF01805; Surp. 2.
CC Transcription regulation; RNA-binding; mRNA splicing; Repeat.

```

```

FT DOMAIN 8 124 DRY CEEERYL.
FT DOMAIN 166 431 2 X REPEATS OF THE SURP MOTIF.
FT REPEAT 166 209 SURP MOTIF 1.
FT REPEAT 391 431 SURP MOTIF 2.
FT DOMAIN 688 749 ARC/LYS/SER-RICH (HIGHLY BASIC).
SQ SEQUENCE 749 AA; 86309 MW; 7F2627E37CD1C390 CRC64;

Query Match 33.3%; Score 62; DB 1; Length 749;
Best Local Similarity 46.2%; Pred. No. 1.9;
Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 EREGSSQOCROEVORLSSCERYL 30
Db 71 KRNSPSEOCPTFAMEEDMCEERYL 96

RESULT 3
2SS_BEREX STANDARD; PRT; 146 AA.
AC P04403; P04402;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S sulfur-rich seed storage protein precursor (Allergen Ber e 1).
GN BZS1 AND BE252.
OS Bertholletia excelsa (Brazil nut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Lecythidaceae; Bertholletia.
OX NCBL_TaxID=3645;
RN [1]
RP SEQUENCE FROM N.A.
RA Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M.;
RT "Cloning and sequence analysis of a cDNA encoding a Brazil nut protein
RT exceptionally rich in methionine.";
RL Plant Mol. Biol. 8:239-250(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Bassener R.;
RL Submitted (DEC-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91370890; PubMed=1840683;
RA Gander E.S., Holmstroem K.O., de Paiva G.R., de Castro L.A.B.,
RA Carneiro M., Grossi de Sa M.F.;
RT "Isolation, characterization and expression of a gene coding for a 2S
RT albumin from Bertholletia excelsa (Brazil nut).";
RL Plant Mol. Biol. 16:437-448(1991).
RN [4]
RP SEQUENCE OF 37-64 AND 70-142.
RX MEDLINE=87004679; PubMed=3758080;
RA Ampe C., van Damme J., de Castro L.A.B., Sampaio M.J.A.M.,
RA van Montagu M., Vandekerckhove J.;
RT "The amino-acid sequence of the 2S sulphur-rich proteins from seeds
RT of Brazil nut (Bertholletia excelsa H.B.K.).";
RL Eur. J. Biochem. 159:597-604(1986).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M17146; AAA33010.1; -.
CC EMBL; X57027; CAA40343.1; -.
CC EMBL; X57028; CAA40344.1; -.

```



```

FT PROPEP 20 29
FT CHAIN 30 148
FT DOMAIN 68 73
SO SOURCE 148 AA; 16792 MW; 327904B4BEC2C16 CRC64;

Query Match
Best Local Similarity 31.28; Score 58; DB 1; Length 148;
Matches 13; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

OY 8 GSSGQCRQEVORKDLSCERYLQ 32
   | |||| | | |||| | : :
Db 34 GGSQCCPQF--RPKSSCKDYWE 56

RESULT 6
CG2S_LUPAN
ID CG2S_LUPAN STANDARD; PRT; 37 AA.
AC P09350;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Conglutin delta-2 small chain.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3871;
RN [1]
RP SEQUENCE.
RC STRAIN-CV. WHITE;
RA Lilly G.G., Ingalls A.S.;
RT "Amino acid sequence of conglutin delta, a sulfur-rich seed protein
RT of lupinus angustifolius L. Sequence homology with the C-II alpha-
RT amylase inhibitor from wheat."
RL FEBS Lett. 195:235-241(1986).
RC -1- SUBUNIT: DIMER OF A SMALL CHAIN AND A LARGE CHAIN LINKED BY TWO
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR PIR; A23617; A23617.
KW Seed.
FT DISULFID 8 8 INTERCHAIN (WITH C-29 OF LARGE CHAIN).
FT DISULFID 20 20 INTERCHAIN (WITH C-17 OR C-18 OF LARGE
FT CHAIN).
FT DOMAIN 29 37 GLU/GLN-RICH.
SQ SEQUENCE 37 AA; 4598 MW; 89784D55A5A1493A CRC64;

Query Match
Best Local Similarity 30.18; Score 56; DB 1; Length 37;
Matches 9; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 9 SSSGQCRQEVORKDLSCERYLQ 32
   | | | | : : : : | | : : |
Db 3 SSSGCRQLQOVNLRHCENHIQ 26

RESULT 7
GLTQ_WHEAT
ID GLTQ_WHEAT STANDARD; PRT; 648 AA.
AC P10387;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit D10 precursor.
GN GLU-D1-2b.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Trilicaceae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-CV. CHEYENNE;
RX MEDLINE=89098419; PubMed=2563152;
RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
RA Malpica-Romero J.M.;
RT "Nucleotide sequences of the two high-molecular-weight glutenin genes
RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
RT Cheyenne."
RL Nucleic Acids Res. 17:461-462(1989).
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISCIO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GOOPGQ AND
CC GOOPGOGGQGYPTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL; X12929; CAA31396.1; -
CC PIR; S04832; S04832.
CC DR HSSP; P01088; 1BFA.
CC DR InterPro; IPR001768; Cereal_tryp_amyl_inh.
CC DR InterPro; IPR001419; Glutenin.
CC DR Pfam; PF00234; tryp_alpha_amyl_1.
CC DR PRINTS; PR00210; GLUTENIN.
CC KW Seed storage protein; Repeat; Multigene family; Signal.
CC FT STGNL 1 21
CC FT CHAIN 22 648 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
CC FT DOMAIN 147 610 D10.
CC FT SEQUENCE 648 AA; 69629 MW; FE98F1D44B9E9AF1 CRC64;

Query Match
Best Local Similarity 29.68; Score 55; DB 1; Length 648;
Matches 11; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

OY 5 EREGSSGQCRQEVORKDLSCERYLQSSRR 37
   | | | | : : : : | | : : |
Db 22 EGEASRQLQCERELQESSLFCRCQVVDQLAGR 54

RESULT 8
GLT3_WHEAT
ID GLT3_WHEAT STANDARD; PRT; 660 AA.
AC P08458;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit 12 precursor.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Trilicaceae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. CHINESE SPRING;
RX MEDLINE=86041882; PubMed=3840588;
RA Thompson R.D., Bartels D., Harberd N.P.;
RT "Nucleotide sequence of a gene from chromosome 1D of wheat encoding a
RT HMW-glutenin subunit."
RL Nucleic Acids Res. 13:6833-6846(1985).
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE

```



```
CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -1- SUBUNIT. DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQ AND
CC GQPGQGGQGYPPS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03041; CAA26847.1; -.
DR PIR; A24266; A24266.
DR HSSP; P01088; 1BFA.
DR InterPro: IPR001768; Cereal tryp-amy1_inh.
DR InterPro: IPR001419; Glutenin.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR PRINTS; PR00210; GLUTENIN.
DR Seed storage protein; Repeat; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 660
FT DOMAIN 125 615 REPEATS.
FT SQUENCE 660 AA; 70868 MW; 2BFD09D8C8FCCEFF CRC64;
SQ
QY Query Match 29.6%; Score 55; DB 1; Length 660;
Best Local Similarity 33.3%; Pred. No. 13;
Matches 11; Conservative 6; Mismatches 16; Indels 0; Gaps 0;
QY 5 EREGSSSQCRQEVORRKLSCRCRYLRSSSR 37
| | | | | | | | | | | | | | | | | | | | |
DB 22 EGASRQLQGERELQESSLEACRQVYDQQLAGR 54
RESULT 9
Y218_HUMAN STANDARD; PRT; 761 AA.
ID Y218_HUMAN
AC Q93075;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative deoxyribonuclease KIAA0218 (EC 3.1.21.-).
GN KIAA0218.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Bone marrow;
RX MEDLINE=97191544; Pubmed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
CC -1- SIMILARITY: BELONGS TO THE TAND DNASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; D86972; BAA13208.1; -.
DR InterPro: IPR001130; TatD.
DR Pfam; PF01026; TatD_DNase; 1.
DR PROSITE; PS01137; TATD_1; 1.
DR PROSITE; PS01090; TATD_2; 1.
DR PROSITE; PS01091; TATD_3; FALSE NEG.
KW Hypothetical protein; Hydrolase; Nuclease.
SQ SEQUENCE 761 AA; 85038 MW; 7B0504FBDD262CA8E CRC64;
QY Query Match 29.6%; Score 55; DB 1; Length 761;
Best Local Similarity 35.1%; Pred. No. 15;
Matches 13; Conservative 5; Mismatches 19; Indels 0; Gaps 0;
QY 1 ESKGERGSSSQCRQEVORRKLSCRCRYLRSSSR 37
| | | | | | | | | | | | | | | | | | | | |
DB 40 QRSASRGSGSPSRKLRKAKEDDVACSRRLSWGSSRR 76
RESULT 10
Y2S4_CAPMA STANDARD; PRT; 100 AA.
ID Y2S4_CAPMA
AC P80353;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin IV, A and B chains (MAB IV) (Sweet protein).
OS Capparis masaiikai (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Capparis.
OX NCBI_Taxid=13395;
RN [1]
RP SEQUENCE.
RC TISSUE-Seed;
RX MEDLINE=94333405; Pubmed=8055976;
RA Nirasawa S., Nishino T., Katahira M., Desugi S., Hu Z., Kurihara Y.;
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.";
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
CC INDUCING ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC InterPro: IPR003612; AAI.
CC InterPro: IPR000617; Napin.
CC Pfam; PF01631; Seedstore_2S; 1.
CC PRINTS; PR00496; NAPIN.
DR PRODOM; PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
DR Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 28
FT NON CONS 28 29
FT CHAIN 28 100
FT DISULFID 4 49
FT DISULFID 17 38
FT DISULFID 39 87
FT DISULFID 51 95
FT SEQUENCE 100 AA; 11928 MW; 7B09673FAB7793CA CRC64;
QY Query Match 28.8%; Score 53.5; DB 1; Length 100;
Best Local Similarity 41.7%; Pred. No. 2.9;
Matches 10; Conservative 8; Mismatches 5; Indels 1; Gaps 1;
QY 14 CROEVOR-KDLSRCRYLRSSSR 36
| | | | | | | | | | | | | | | | | | | | |
DB 4 CRRQFOOHQHLRACQRYLRRAQR 27
RESULT 11
```

```

25S3_CAPMA
ID 25S3_CAPMA STANDARD: PRT: 104 AA.
AC P80352:
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Mabinlin II, A and B chains (MAB III) (Sweet protein).
OS Capparis masakal (Mablinang).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC eucosids II: Brassicales: Brassicaceae: Capparids.
OX NCBI_TaxId=13395;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed:
RX MEDLINE=94333405; PubMed=8055976;
RA Nirasawa S., Nishino T., Katahira M., Uesugi S., Hu Z., Kurihara Y.;
RT Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.*;
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
CC INDUCING ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR Interpro: IPR003612; AAI.
DR Interpro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR ProDom: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein: Albumin; Sweet-taste.
FT CHAIN 1 32 MABINLIN III, A CHAIN.
FT NON-CONS 32 33
FT CHAIN 33 104 MABINLIN III, B CHAIN.
FT DISULFID 4 53
FT DISULFID 17 42
FT DISULFID 43 91
FT DISULFID 55 99
SQ SEQUENCE 104 AA: 12284 MW: 102EB3F5F24AD3D0 CRC64;

Query Match 28.8%; Score 53.5; DB 1; Length 104;
Best Local Similarity 41.7%; Pred. No. 3;
Matches 10; Conservative 8; Mismatches 5; Indels 1; Gaps 1;

OY 14 CROEVOR-KDLSCEERYLRQSSR 36
DB 4 CROFOOHOLRACORRYLRRAQR 27.

RESULT 12
GLT4_WHEAT
ID GLT4_WHEAT STANDARD: PRT: 838 AA.
AC P08489;
DT 01-AUG-1988 (rel. 08, Created)
DT 01-AUG-1988 (rel. 08, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE Glutelin, high molecular weight subunit FW212 precursor.
OS Triticum aestivum (Wheat).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae; Pooidae;
OC Triticaceae; Triticum.
OX NCBI_TaxId=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. YAMHILL;
RX MEDLINE=86093674; PubMed=3001648;
RA Sugiyama T., Rafalski A., Peterson D., Soll D.G.;
RT "A wheat HMW glutenin subunit gene reveals a highly repeated
RT structure.";
RL Nucleic Acids Res. 13:8729-8737(1985).

```

```

CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPPQ AND
CC GQPPGQGGQGYTPS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: X03346; CAA27052.1; -.
DR PIR: A24107; EEMTHW.
DR Interpro: IPR001419; Glutenin.
DR PRINTS: PR00210; GLUTENIN.
KW Seed storage protein; Repeat; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 838 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT DOMAIN 130 799 REPEATS.
FT FT FW212.
SQ SEQUENCE 838 AA: 89174 MW: 71D715B7BDF0722D CRC64;

Query Match 28.8%; Score 53.5; DB 1; Length 838;
Best Local Similarity 35.5%; Pred. No. 25;
Matches 11; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

OY 5 EREGGSSQOCR--QEVQRKDLSCEERYLRQ 32
DB 22 EGEASELOQCERELQELQERELKACQVMDQ 52

RESULT 13
GLT5_WHEAT
ID GLT5_WHEAT STANDARD: PRT: 839 AA.
AC P10388;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE Glutelin, high molecular weight subunit DX5 precursor.
OS GLU-1D-ID OR GLU-D1-1B.
OC Triticum aestivum (Wheat).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae; Pooidae;
OC Triticaceae; Triticum.
OX NCBI_TaxId=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHEYENNE;
RX MEDLINE=89098419; PubMed=2563152;
RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
RA Malpica-Romero J.M.;
RT "Nucleotide sequences of the two high-molecular-weight glutenin genes
RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
RT Cheyenne.";
RL Nucleic Acids Res. 17:461-462(1989).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=CV. CHEYENNE;
RA Anderson O.D.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE

```

```

CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQDPGQ AND
CC GQDPGGGQGGGPPTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X12928; CA81395.1; -.
DR PIR: S02262; S02262.
DR InterPro: IPR001419; Glutinin.
DR PRINTS; PR00210; GLUTENIN.
KM Seed storage protein; Repeat; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 839
FT DOMAIN 131 801 REPEATS.
FT SEQUENCE 839 AA; 89359 MW; 0F14E1106D552643 CRC64;

Query Match 28.8%; Score 53.5; DB 1; Length 839;
Best Local Similarity 35.5%; Pred. No. 25;
Matches 11; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

OY 5 EREGSSSQCR---OEVOQRKDLSCERYLRQ 32
DB 22 EGEASEQACERELQELQERELKACQYMDQ 52

RESULT 14
RNL_AERRY STANDARD: PRT; 215 AA.
AC 007465;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease precursor (EC 3.1.27.-).
OS Aeromonas hydrophila.
CC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
CC Aeromonas.
OX NCBI_TaxID=644;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH1133;
RX MEDLINE=93285983; PubMed=7685334;
RA Favre D., Ngai P.K., Timmis K.N.;
RT "Relatedness of a periplasmic, broad-specificity RNase from Aeromonas
RT hydrophila to RNase I of Escherichia coli and to a family of
RT eukaryotic RNases."
RL J. Bacteriol. 175:3710-3722(1993).
CC -1- FUNCTION: ONE OF THE FEW RNASES THAT CLEAVES THE PHOSPHODIESTER
CC BOND BETWEEN ANY TWO NUCLEOTIDE. SHOWS A PREFERENCE FOR ADENYLIC
CC ACID.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC AND CYTOPLASMIC. AN RNASE I-LIKE
CC FORM (PERIPLASMIC) AND RNASE I*-LIKE FORM (CYTOPLASMIC) APPEAR TO
CC BE ISOFORMS APPARENTLY ENCODED BY THE SAME GENE. THE CYTOPLASMIC
CC FORM IS LESS ACTIVE TOWARDS NATURAL POLYMER RNA.
CC -1- SIMILARITY: BELONGS TO THE RNASE T2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: X67054; CA847438.1; -.
DR PIR: A47118; A47118.
DR PIR: S23227; S23227.
DR InterPro: IPR001568; RNase-T2.
DR Pfam: PF00445; ribonuclease_T2; 1.
DR PROSITE: PS00530; RNASE_T2.1; 1.
DR PROSITE: PS00531; RNASE_T2.2; 1.
KM Hydroxylase, Nuclease; Endonuclease; Periplasmic; Signal.
FT SIGNAL 1 22
FT CHAIN 23 215
FT ACT_SITE 62 62 BY SIMILARITY.
FT ACT_SITE 102 102 BY SIMILARITY.
FT ACT_SITE 106 106 BY SIMILARITY.
FT ACT_SITE 109 109 BY SIMILARITY.
FT DISULFID 76
FT SEQUENCE 215 AA; 24410 MW; 4442BCE5B0F67203 CRC64;

Query Match 28.5%; Score 53; DB 1; Length 215;
Best Local Similarity 38.1%; Pred. No. 7.3;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 6 EREGSSSQCRQEVQRKDLSSC 26
DB 184 KEGTDAVTCSDMKRKLPC 204

RESULT 15
GSH1_BUCAI STANDARD: PRT; 518 AA.
ID GSH1_BUCAI
AC P57485;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamate--cysteine ligase (EC 6.3.2.2) (Gamma-glutamylcysteine
DE synthetase) (Gamma-ECS) (GCS).
GN GSHA OR BU407.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
CC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TOKYO.1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sasaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + L-cysteine -> ADP +
CC phosphate + gamma-L-glutamyl-L-cysteine.
CC -1- PATHWAY: FIRST STEP IN GLUTATHIONE BIOSYNTHESIS.
CC -1- SIMILARITY: STRONG, TO E.COLI GSH1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001119; BAB13108.1; -.
DR Glutathione biosynthesis; ligase; Complete proteome.
KM SEQUENCE 518 AA; 60979 MW; AC659F1E2C50BD4 CRC64;

Query Match 28.5%; Score 53; DB 1; Length 518;
Best Local Similarity 40.7%; Pred. No. 18;
Matches 11; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

OY 5 EREGSSSQCRQEVQRKDLSCERYLR 31
DB 490 EGEVIRSHQKRIEREDILSFEEYIR 516

```

Search completed: September 9, 2002, 12:46:06
Job time: 308 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:40:43 : Search time 77.42 Seconds
(without alignments)
82.676 Million cell updates/sec

Title: US-09-913-351-2

Perfect score: 186

Sequence: 1 ESKGEREGSSQOCROEVQRKDLSCERYLRQSSRR 37

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186	100.0	41	10	Q9S874
2	79	42.5	153	10	Q9AUD1
3	75.5	40.6	141	10	Q39649
4	74	39.8	323	10	Q39928
5	62	33.3	747	5	Q17484
6	61.5	33.1	146	10	Q9LRC2
7	61.5	33.1	160	10	Q9SOH1
8	61	32.8	572	10	Q9A064
9	60.5	32.5	139	10	P93198
10	59	31.7	471	2	Q9EXV9
11	59	31.7	765	12	Q9JGS5
12	58	31.2	463	10	Q94I09
13	58	31.2	742	10	Q94I10
14	58	31.2	754	10	Q94I12
15	58	31.2	754	10	Q94IK9
16	58	31.2	754	10	Q94IK9

17	58	31.2	754	10	Q93WF0	Q93WF0 secale cere
18	58	31.2	766	10	Q9SDM3	Q9SDM3 triticum ae
19	58	31.2	766	10	Q94IL3	Q94IL3 secale cere
20	58	31.2	811	10	Q94I17	Q94I17 triticum ae
21	57	30.6	153	10	Q99235	Q99235 lupinus ang
22	57	30.6	392	5	Q95R97	Q95R97 drosophila
23	57	30.6	699	10	Q9ZU69	Q9ZU69 arabidopsis
24	57	30.6	904	5	Q9VC33	Q9VC33 drosophila
25	56	30.1	141	10	Q9LE34	Q9LE34 hordeum vul
26	56	30.1	141	10	Q9M3V3	Q9M3V3 hordeum vul
27	56	30.1	142	10	Q9M3V6	Q9M3V6 hordeum vul
28	56	30.1	142	10	Q9M3V5	Q9M3V5 hordeum vul
29	56	30.1	142	10	Q9M3V4	Q9M3V4 hordeum vul
30	56	30.1	147	10	Q9FS19	Q9FS19 hordeum vul
31	56	30.1	147	10	Q9LEH8	Q9LEH8 hordeum vul
32	56	30.1	765	12	Q9JGS2	Q9JGS2 triticum ae
33	55	29.6	156	10	Q9M4D9	Q9M4D9 triticum mo
34	55	29.6	156	10	Q94IR0	Q94IR0 arachis hyp
35	55	29.6	179	10	Q9M5N3	Q9M5N3 aegilops ve
36	55	29.6	241	5	Q23533	Q23533 caenorhabdi
37	55	29.6	648	10	Q38767	Q38767 aegilops ta
38	55	29.6	654	10	Q93XR5	Q93XR5 aegilops cy
39	55	29.6	705	10	Q03871	Q03871 triticum ae
40	55	29.6	707	10	Q94IL6	Q94IL6 secale cere
41	55	29.6	713	10	Q9SDM2	Q9SDM2 triticum ae
42	55	29.6	713	10	Q94IT1	Q94IT1 secale cere
43	55	29.6	713	10	Q94IT8	Q94IT8 triticum ae
44	55	29.6	713	10	Q93WM1	Q93WM1 secale cere
45	55	29.6	720	10	Q94I06	Q94I06 triticum ae

ALIGNMENTS

RESULT 1					
Q9S874	PRELIMINARY:	PRT:	41 AA.		
AC Q9S874:					
DT 01-MAY-2000 (TREMBlrel. 13, Created)					
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)					
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)					
DE 4.7 KDa NAPIN-LIKE PROTEIN SMALL CHAIN (FRAGMENT).					
OS Ricinus communis (Castor bean).					
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;					
OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.					
OX NCBI_TaxID=3988;					
RN [1]					
RP SEQUENCE.					
RX MEDLINE=97135090; PubMed=8980648;					
RA Neumann G.M., Condron R., Polya G.M.;					
RT "Purification and sequencing of napin-like protein small and large					
RT chains from Momordica charantia and Ricinus communis seeds and					
RT determination of sites phosphorylated by plant Ca(2+)-dependent					
RT protein kinase."; Acta 1298:223-240(1996).					
RL Biochim. Biophys. Acta 1298:223-240(1996).					
SQ SEQUENCE 41 AA: 4732 MW; 201FE2938F33173 CRC64;					
Query Match	100.0%;	Score 186;	DB 10;	Length 41;	
Best local similarity	100.0%;	Pred. No. 6.3e-19;			
Matches 37;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 1 ESKGEREGSSQOCROEVQRKDLSCERYLRQSSRR 37					
DB 1 ESKGEREGSSQOCROEVQRKDLSCERYLRQSSRR 37					
RESULT 2					
Q9AUD1	PRELIMINARY:	PRT:	153 AA.		
AC Q9AUD1:					
DT 01-JUN-2001 (TREMBlrel. 17, Created)					

RA None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 RL investigating biology. The *C. elegans* Sequencing Consortium.";
 RN Science 282:2012-2018(1998).
 RP [2]
 RC SEQUENCE FROM N.A.
 RA STRAIN-BRISTOL N2;
 RT "The sequence of *C. elegans* cosmid B0336.";
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RC SEQUENCE FROM N.A.
 RA STRAIN-BRISTOL N2;
 RL Waterston R.;
 DR Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U32305; AAK18860.1; -
 DR Interpro: IPR000061; Surp.
 DR Pfam: PF01805; Surp; 2.
 SQ SEQUENCE 747 AA; 86068 MW; EB0A50EBF659B99E CRC64;

Query Match 33.3%; Score 62; DB 5; Length 747;
 Best Local Similarity 46.2%; Pred. No. 2.1;
 Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 5 EREGSSQCCROEVQRKDLSSCERYL 30
 DB 71 KRNGSPSCPTAMEEDMCEERYL 96

RESULT 6
 Q9LRC2 PRELIMINARY; PRT; 146 AA.
 AC Q9LRC2;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 2S ALBUMIN
 OS Bertholletia excelisa (Brazil nut).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Ericales; Lecythidaceae; Bertholletia.
 OX NCBI_TaxID=3645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamauchi D.;
 RT "Brazil nut 2S albumin was synthesized in a transgenic French bean
 RT seed with a promoter of the gene for canavalin, 7S globulin from
 RT Canavalia gladiata.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB044391; BAA96554.1; -
 DR HSSP; P01087; IB10.
 DR Interpro: IPR003612; AAI.
 DR Interpro: IPR001768; Cereal_tryp_amy1_inh.
 DR Interpro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_amy1. 1.
 DR PRINTS; PRO0808; AMLASEIHBBR.
 DR PRINTS; PRO0496; NAPIIN.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 146 AA; 16910 MW; 4A69A196E6EC7096 CRC64;

Query Match 33.1%; Score 61.5; DB 10; Length 146;
 Best Local Similarity 40.7%; Pred. No. 0.49;
 Matches 11; Conservative 8; Mismatches 7; Indels 1; Gaps 1;

QY 7 EGSSSQCCROEVQRKDLSSCERYL 32
 DB 33 EENQECRCRQMOQOMLSHCRTMRO 59

RESULT 7
 Q9SOH1

ID Q9SOH1 PRELIMINARY; PRT; 160 AA.
 AC Q9SOH1;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ALLERGEN.
 GN ARA H 7.
 OS Arachis hypogaea (peanut).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
 OC Arachis.
 OX NCBI_TaxID=3818;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VIRGINIA; TISSUE=SEED;
 RX MEDLINE=99406463; Pubmed=10474031;
 RA Kleber-Janke T., Crameri R., Appenzeller U., Schlaak M., Becker W.M.;
 RT "selective cloning of peanut allergens, including profilin and 2S
 RT albumins, by phage display technology.";
 RL Int. Arch. Allergy Immunol. 119:265-274(1999).
 DR EMBL; AF091737; AAD56719.1; -
 DR Interpro: IPR003612; AAI.
 DR Interpro: IPR001768; Cereal_tryp_amy1_inh.
 DR Pfam: PF00234; tryp_alpha_amy1. 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 160 AA; 18417 MW; 9F94CEB68808D4C CRC64;

Query Match 33.1%; Score 61.5; DB 10; Length 160;
 Best Local Similarity 35.3%; Pred. No. 0.53;
 Matches 12; Conservative 11; Mismatches 8; Indels 3; Gaps 1;

QY 2 SKGERGSSSQ--QCRQEVQRKDLSSCERYL 32
 DB 30 SRGSRMDAPSRGDDQCRQLGRANLRPCHEHMR 63

RESULT 8
 Q9AU64 PRELIMINARY; PRT; 572 AA.
 AC Q9AU64;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 7S GLOBULIN.
 GN GLO7A.
 OS Elaeis guineensis var. tenera (oil palm).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Arecaeae; Arecoidae;
 OC Coccaeae; Elaeidinae; Elaeis.
 OX NCBI_TaxID=51953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Morcillo F., Hartmann C., Duval Y., Tregear J.;
 RT "regulation of 7S globulin gene expression in zygotic and somatic
 RT embryos of oil palm.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF250228; AAK28402.1; -
 DR HSSP; P02853; 2PHL.
 DR Interpro: IPR001113; Seedstore_7s.
 DR Pfam; PF00546; Seedstore_7s. 1.
 DR Pfam; PF02808; Seedstore_7s-C; 1.
 SQ SEQUENCE 572 AA; 66364 MW; 606A071866A6FB3B CRC64;

Query Match 32.8%; Score 61; DB 10; Length 572;
 Best Local Similarity 36.4%; Pred. No. 2.3;
 Matches 12; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 5 EREGSSQCCROEVQRKDLSSCERYLRSSRR 37
 DB 74 ERKGGEGEGRGRREPERKRLDECRCREGRQARR 106

RESULT 9
ID 09JGS8 PRELIMINARY; PRT; 765 AA.
AC 09JGS8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PORFL.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20417334; PubMed=10963344;
RA Tanaka Y., Orito E., Ohno T., Nakano T., Hayashi K., Kato T.,
Mukaido M., Iida S., Mizokami M.;
RT "Identification of a 23kda protein encoded by putative open reading
frame 2 of TT virus (TTV) genotype I different from the other
genotypes.";
RL Arch. Virol. 145:1385-1398(2000).
DR EMBL; AB030487; BAA90406.1; -;
DR InterPro: IPR004219; TT_ORF1.
DR Pfam: PF02956; TT_ORF1.1.
SQ SEQUENCE 765 AA; 90225 MW; E845A26B6D9D707 CRC64;

Query Match 32.8%; Score 61; DB 12; Length 765;
Best Local Similarity 46.9%; Pred. No. 3;
Matches 15; Conservative 7; Mismatches 8; Indels 2; Gaps 1;

QY 1 ESKGEREGSSQOCROEVORRDLSCERYLRQ 32
Db 704 ESQKEHGTLSQIREQVQOKLGL--RQLARE 733

RESULT 10
ID P93198 PRELIMINARY; PRT; 139 AA.
AC P93198;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALBUMIN SEED STORAGE PROTEIN (FRAGMENT).
OS Juglans regia (English walnut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fagales; Juglandaceae; Juglans.
OX NCBI_TaxID=51240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNLAND;
RA Teuber S.S., Dandekar A.M., Peterson W.R.;
RT "Juglans regia 2S albumin seed storage protein precursor.";
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U66866; AAB41308.1; -;
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal.tryp-amyL_inh.
DR InterPro: IPR000480; Glutelin.
DR Pfam: PF00234; tryp_alpha-amyL.1.
DR PRINTS; PR00211; GLUTELIN.
DR SMART; SM00499; AAI.1.
FT NON_TER 1
SQ SEQUENCE 139 AA; 16373 MW; 02D0E55E67164F23 CRC64;

Query Match 32.5%; Score 60.5; DB 10; Length 139;
Best Local Similarity 50.0%; Pred. No. 0.64;
Matches 11; Conservative 7; Mismatches 3; Indels 1; Gaps 1;

QY 14 CROEVOR-KDLSCERYLRQSS 34
15:::11:::1:1:11111

Db 39 CREQIQRQNLNHCQYLRQSS 60

RESULT 11
ID 09EXV9 PRELIMINARY; PRT; 471 AA.
AC 09EXV9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GDP-MANNOSE PYROPHOSPHORYLASE.
GN MANC.
OS Salmonella enterica subsp. enterica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=59201;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M252;
RX MEDLINE=21135136; PubMed=11238967;
RA Jensen S.O., Reeves P.R.;
RT "Molecular evolution of the GDP-mannose pathway genes (manB and manC)
in Salmonella enterica.";
RL Microbiology 147:599-610(2001).
DR EMBL; AF012201; AAG41707.1; -;
DR InterPro: IPR001538; MannoseP_isomer.
DR InterPro: IPR001825; NTP_transferase.
DR Pfam: PF01050; MannoseP_isomer.1.
DR Pfam: PF00483; NTP_transferase.1.
DR ProDom; PD002664; MannoseP_isomer.1.
SQ SEQUENCE 471 AA; 52962 MW; E3CB9A2CAF636A90 CRC64;

Query Match 31.7%; Score 59; DB 2; Length 471;
Best Local Similarity 38.7%; Pred. No. 3.5;
Matches 12; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 3 KGEREGSSQOCROEVORRDLSCERYLRQ 33
Db 157 KGERVNDSSQFVESFVEKPDIEFATKDYLRQN 187

RESULT 12
ID 09JGS5 PRELIMINARY; PRT; 765 AA.
AC 09JGS5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PORFL.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20417334; PubMed=10963344;
RA Tanaka Y., Orito E., Ohno T., Nakano T., Hayashi K., Kato T.,
Mukaido M., Iida S., Mizokami M.;
RT "Identification of a 23kda protein encoded by putative open reading
frame 2 of TT virus (TTV) genotype I different from the other
genotypes.";
RL Arch. Virol. 145:1385-1398(2000).
DR EMBL; AB030488; BAA90409.1; -;
DR InterPro: IPR004219; TT_ORF1.
DR Pfam: PF02956; TT_ORF1.1.
SQ SEQUENCE 765 AA; 90275 MW; 707029EBA5829B5F CRC64;

Query Match 31.7%; Score 59; DB 12; Length 765;
Best Local Similarity 43.8%; Pred. No. 5.7;
Matches 14; Conservative 9; Mismatches 7; Indels 2; Gaps 1;

QY 1 ESKGEREGSSQOCROEVORRDLSCERYLRQ 32

Db 704 ESQKQEGTISQQLREQLQKLLG--RQLE 733

RESULT 13
0941J9 PRELIMINARY: PRT; 462 AA.

AC 0941J9; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE HIGH MOLECULAR WEIGHT GLUTENIN SUBUNIT X PRECURSOR.
GN GLU-1R.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Trilicaceae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. S-149;
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. S-149;
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314772; CAC40684.1; -.
KW SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT X.
FT SEQUENCE 22 462 79C3BD615F91BF12 CRC64;
SQ

Query Match 31.2%; Score 58; DB 10; Length 462;
Best Local Similarity 35.7%; Pred. No. 4.8;
Matches 10; Conservative 7; Mismatches 11; Indels 0; Gaps 0;
OY 5 EREGSSQGCROEYQRDLSSCERYLRQ 32
Db 22 EGEASGQLCCERLELEACRQIVDQ 49

RESULT 14
0941I0 PRELIMINARY: PRT; 743 AA.
AC 0941I0; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE HIGH MOLECULAR WEIGHT GLUTENIN SUBUNIT X PRECURSOR.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Trilicaceae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314776; CAC40677.1; -.
KW SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT X.
FT SEQUENCE 22 743 89BAD795A3DC91BB CRC64;
SQ

Query Match 31.2%; Score 58; DB 10; Length 743;
Best Local Similarity 35.7%; Pred. No. 7.7;
Matches 10; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

OY 5 EREGSSQGCROEYQRDLSSCERYLRQ 32
Db 22 EGEASGQLCCERLELEACRQIVDQ 49

RESULT 15
0941I2 PRELIMINARY: PRT; 754 AA.

AC 0941I2; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE HIGH MOLECULAR WEIGHT GLUTENIN SUBUNIT X PRECURSOR.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Trilicaceae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314773; CAC40674.1; -.
KW SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT X.
FT SEQUENCE 22 754 89BAD795A3DC91BB CRC64;
SQ

Query Match 31.2%; Score 58; DB 10; Length 754;
Best Local Similarity 35.7%; Pred. No. 7.8;
Matches 10; Conservative 7; Mismatches 11; Indels 0; Gaps 0;
OY 5 EREGSSQGCROEYQRDLSSCERYLRQ 32
Db 22 EGEASGQLCCERLELEACRQIVDQ 49

Search completed: September 9, 2002, 12:45:36
Job time: 293 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:43:30 ; Search time 45.25 seconds
(without alignments)
144.399 Million cell updates/sec

Title: US-09-913-351-3
Perfect score: 357
Sequence: 1 QOSESQQLQCCNQVKQVRD.....AQRAGEIVSQCVCRCMQRTR 68

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357	100.0	258	1 RZCS	2S seed storage pr
2	132.5	37.1	223	2 S38887	2S albumin - commo
3	125	35.0	295	2 S01062	2S seed storage pr
4	124	34.7	70	2 A59346	2S albumin prote
5	122	34.2	141	2 T10257	2S albumin precu
6	113.5	31.8	154	2 A14947	2S albumin - Brazi
7	110	30.8	164	1 NMMU1	2S albumin 1 precu
8	107.5	30.1	165	2 T09252	seed storage prote
9	104.5	29.3	100	2 S48180	mablinin IV - Yun
10	104.5	29.3	104	2 S48176	mablinin IRI - Yun
11	103.5	29.0	104	2 S48176	mablinin IRI - Yun
12	102	28.6	153	2 A33090	conglutin delta pr
13	101	28.3	164	1 NMMU3	2S albumin 3 precu
14	98	27.5	80	2 B23617	conglutin delta-2
15	95.5	26.8	162	2 T08010	2S seed storage pr
16	95	26.6	162	2 S49259	albumin 4 - easter
17	91.5	25.6	124	2 A37931	albumin - Swedish tu
18	91	25.5	178	1 NMRP2	napin 2 precursor
19	91	25.5	178	2 A25997	napin precursor (n
20	91	25.5	178	2 S25130	2S storage protein
21	91	25.5	178	2 S25127	2S storage protein
22	91	25.5	180	2 S10018	napin (clone BqNA
23	91	25.5	180	2 S52025	napin (clones BNMN
24	90.5	25.4	92	2 S35592	major allergen lar
25	90	25.2	173	2 T08011	2S seed storage pr
26	89	24.9	162	2 T08013	2S seed storage pr
27	89	24.9	170	1 NMMU2	2S albumin 2 precu
28	88	24.6	139	2 T09878	albumin 2S storage
29	88	24.6	172	2 S18871	2S-like storage pr

30	88	24.6	285	2 S20853	glutenin low molec
31	87.5	24.5	152	2 PS0427	napin AHI precuro
32	87	24.4	186	2 A29802	napin precursor (g
33	86.5	24.2	145	2 PC1247	Sin a I allergen 2
34	86.5	24.2	155	2 JC5379	mablinin II precu
35	86	24.1	127	2 S65447	allergen Sin a I -
36	86	24.1	178	2 S07828	napin B - rape
37	85.5	23.9	72	2 S28843	mablinin II chain
38	85.5	23.9	146	2 S14946	2S seed storage pr
39	85.5	23.9	170	2 T08012	2S seed storage pr
40	85.5	23.9	174	2 PS0425	napin B43 precuro
41	85.5	23.9	319	2 A22364	alpha/beta-gliadin
42	85	23.8	155	2 PS0426	napin B69 precuro
43	84.5	23.7	148	2 A46514	puroindoline-b pre
44	84.5	23.7	166	1 NMMU4	2S albumin 4 precu
45	84.5	23.7	320	2 E22364	alpha/beta-gliadin

ALIGNMENTS

RESULT 1
RZCS
2S seed storage protein precursor - castor bean
N:Alternate names: 2S albumin precursor
C:Species: Ricinus communis (castor bean)
C>Date: 14-Nov-1983 #sequence:revision 08-Feb-1996 #text_change 18-Jun-1999
C:Accession: S11499; S11500; S11501; S11502; S27221; A01328; S27222
R:Irwin, S.D.; Lord, J.M.
Nucleic Acids Res. 18, 5890, 1990
A:Title: Nucleotide sequence of a Ricinus communis 2S albumin precursor gene.
A:Reference number: S11499; MID:91016940
A:Accession: S11499
A:Molecule type: DNA
A:Residues: 1-258 <IR4>
A:Cross-references: EMBL:X54158; NID:921067; PIDN:CAA38097.1; PID:921068
A:Note: the authors translated the codon CAC for residue 14 as Phe, CCA for residue 7
R:Irwin, S.D.; Keen, J.N.; Findlay, J.B.C.; Lord, J.M.
Mol. Gen. Genet. 222, 400-408, 1990
A:Title: The Ricinus communis 2S albumin precursor: a single preproprotein may be pro
A:Reference number: S11500; MID:91109729
A:Accession: S11500
A:Molecule type: mRNA
A:Residues: 1-13, 'F', 15-73, 'T', 75-258 <IR2>
A:Experimental source: clone 1494
A:Accession: S11501
A:Molecule type: mRNA
A:Residues: 'M', 4, 'L', 7-13, 'F', 15-21 <IRF>
A:Experimental source: clone 10a12
A:Accession: S27221
A:Residues: 'SFAIVYF', 15-73, 'T', 75-258 <IR3>
A:Experimental source: clone 8g8
A:Accession: S11502
A:Molecule type: mRNA
A:Residues: 'M', 4, 'L', 7-13, 'F', 15-21 <IRF>
A:Experimental source: clone 10a12
A:Accession: S27221
A:Residues: 'X', 37, 'X', 39-45, 'X', 158-161, 'X', 163-174, 'X' <IRA>
R:Shatleff, F.S.; Li, S.S.L.
J. Biol. Chem. 257, 14753-14759, 1982
A:Title: Amino acid sequence of small and large subunits of seed storage protein from
A:Reference number: A93357; MID:83082772
A:Accession: A01328
A:Molecule type: protein
A:Residues: 157-190, 194-221, 'Q', 223-225, 230-233, 'N', 235-254, 'Q', 256-258 <SHA>
A:Note: 230-Ser was also found
A:Note: there is considerable similarity between residues 181-231 of this protein and
R:Odani, S.; Koide, T.; Ono, T.; Ohnishi, K.
Biochem. J. 213, 543-545, 1983
A:Title: Structural relationship between barley (Hordeum vulgare) trypsin inhibitor a
A:Reference number: A90322; MID:83308577
A:Contents: annotation
A:Note: this protein is homologous with trypsin inhibitor from barley
C:Complex: consists of two chains linked by two disulfide bonds involving Cys-162 and
C:Superfamily: 2S seed storage protein precursor

C:Keywords: pyroglutamic acid; seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-35/Domain: signal sequence #status predicted <PRO>
F:36-72/Product: probable 2S seed storage protein small chain large 2 #status experimental <SC>
F:87-156/Product: probable 2S seed storage protein small chain #status experimental <SM>
F:157-190/Product: 2S seed storage protein small chain #status experimental <SM>
F:194-255/Product: 2S seed storage protein large chain #status experimental <LKG>
F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 100.0%; Score 357; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.3e-29;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 000ESQ0LQCCNQVQVDECCCEAIXIYIAEDQIQGOLHGESESVAAQRAEIVSSCG 60
|||||
DB 87 000ESQ0LQCCNQVQVDECCCEAIXIYIAEDQIQGOLHGESESVAAQRAEIVSSCG 146

OY 61 VRCMRQTR 68
|||||
DB 147 VRCMRQTR 154

RESULT 2
S3887
2S albumin - common sunflower (fragment)
C:Species: Helianthus annuus (common sunflower)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-May-2000
C:Accession: S38887
R:Thyots, P.J.E.; Milllichip, M.; Stobart, A.K.; Griffiths, W.T.; Napier, J.A.; Shewry, F.
submitted to the EMBL Data Library, November 1993
A:Reference number: S38887
A:Accession: S38887
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-323 <THO>
A:Cross-references: EMBL:X76101; NID:9429181; PID:9429182
C:Superfamily: gliadin

Query Match 37.1%; Score 132.5; DB 2; Length 323;
Best Local Similarity 40.6%; Pred. No. 2.3e-06;
Matches 26; Conservative 14; Mismatches 19; Indels 5; Gaps 1;

OY 1 000ESQ0LQCCNQVQVDECCCEAIXIYIAEDQIQGOLHGESESVAAQRAEIVSSCG 55
|||||
DB 249 RQ0EQ0G1QCCNQVQVDECCCEAIXIYIAEDQIQGOLHGESESVAAQRAEIVSSCG 308

OY 56 VSSC 59
:|
DB 309 PNOQ 312

RESULT 3
S01062
2S seed storage protein precursor (clone Hag5) - common sunflower
N:Alternate names: 2S albumin storage protein
C:Species: Helianthus annuus (common sunflower)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 21-Jul-2000
C:Accession: S01062
R:Allen, R.D.; Cohen, E.A.; Vonder Haar, R.A.; Adams, C.A.; Ma, D.P.; Nessler, C.L.; The
Mol. Gen. Genet. 210, 211-218, 1987
A:Title: Sequence and expression of a gene encoding an albumin storage protein in sunfl
A:Reference number: S01062; MUID:86142538
A:Accession: S01062
A:Molecule type: DNA
A:Residues: 1-295 <ALL>
A:Cross-references: EMBL:X06410; NID:918799; PIDN:CAA29699.1; PID:918800
A>Note: part of this sequence, including the amino end of the mature protein, was confi
C:Genetics:
A:Introns: 192/2
C:Superfamily: gliadin
F:1-20/Domain: signal sequence #status predicted <SIG>

F:162-295/Product: 2S seed storage protein #status experimental <MAT>

Query Match 35.0%; Score 125; DB 2; Length 295;
Best Local Similarity 31.1%; Pred. No. 1.2e-05;
Matches 23; Conservative 22; Mismatches 19; Indels 10; Gaps 2;

OY 2 000ESQ0LQCCNQVQVDECCCEAIXIYIAEDQIQGOLHGESESVAAQRAEIVSSCG 51
:|||||
DB 74 RQ0EQ0LQCCNQVQVDECCCEAIXIYIAEDQIQGOLHGESESVAAQRAEIVSSCG 133

OY 52 AGEIVSSCGVCMR 65
:|
DB 134 AQILPVMCMQSR 147

RESULT 4
A59346
seed storage protein Lec2SAl large chain [imported] - tomato (fragment)
N:Alternate names: 2S albumin large subunit
C:Species: Lycopersicon esculentum (tomato)
C:Date: 31-Dec-2001 #sequence_revision 31-Dec-2001 #text_change 31-Dec-2001
C:Accession: A59346
R:Oguri, S.
submitted to the Protein Sequence Database, September 2000
A:Reference number: A59346
A:Accession: A59346
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-70 <OGU>
A:Experimental source: cv. cherry; seed
A>Note: cross-reactive with anti-tomato lectin antiserum; one of four isoforms in see
C:Complex: heterodimer of large and small (see PIR:S59346) chains, disulfide linked
C:Keywords: heterodimer; pyroglutamic acid; seed; storage protein
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 34.7%; Score 124; DB 2; Length 70;
Best Local Similarity 32.9%; Pred. No. 4.4e-06;
Matches 23; Conservative 20; Mismatches 23; Indels 4; Gaps 2;

OY 2 000ESQ0LQCCNQVQVDECCCEAIXIYIAEDQIQGOLHGESESVAAQRAEIVSSCG 60
|||
DB 1 001PRDVQCCNQVQVDECCCEAIXIYIAEDQIQGOLHGESESVAAQRAEIVSSCG 60

OY 61 V---RCMRQF 67
:|
DB 61 LSPRCETQ 70

RESULT 5
T10257
2S albumin precursor - cucurbit
N:Alternate names: prepro2S albumin
C:Species: Cucurbita sp. (cucurbit)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10257; S19323
R:Hara-Nishimura, I.; Takeuchi, Y.; Inoue, K.; Nishimura, M.
Plant J. 4, 793-800, 1993
A:Title: Vesicle transport and processing of the precursor to 2S albumin in pumpkin.
A:Reference number: T17000; MUID:94100993
A:Accession: T10257
A:Status: translated from GR/EMBL/DDBJ
A:Molecule type: mRNA; protein
A:Residues: 1-141 <HAR1>
A:Cross-references: EMBL:D16560; NID:9459404; PIDN:BAA03993.1; PID:9459405
A:Experimental source: seed; storage deposition stage; cotyledon
A>Note: soluble seed protein
R:Hara-Nishimura, I.; Inoue, K.; Nishimura, M.
FEBS Lett. 294, 89-93, 1991
A:Title: A unique vacuolar processing enzyme responsible for conversion of several pr
A:Reference number: S19323; MUID:92077151
A:Accession: S19323

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 26-45;65-84 <HAR2>
C:Keywords: seed; storage protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-35/Domain: propeptide #status experimental <PRO>
F:36-141/Product: 2S albumin #status experimental <MAT>

Query Match 34.2%; Score 122; DB 2; Length 141;
Best Local Similarity 32.8%; Pred. No. 1.3e-05;
Matches 21; Conservative 18; Mismatches 23; Indels 2; Gaps 1;

QY 2 QOESQQLQCCNOKVQVRDECCQCAIKYIAEDQIQGQLHGESERVAQRAGEIVSSCGV 61
DB 77 RREGSGFECCRELKNVDECRCDLIERARE--EORARQGEGRQMLQKARNLPSCMGI 134
QY 62 RCMR 65
DB 135 RPQR 138

RESULT 6
2S albumin - Brazil nut
C:Species: Bertholletia excelsa (Brazil nut)
C:Date: 13-Jan-1995 #sequence, revision 13-Jan-1995 #text, change 21-Jul-2000
C:Accession: S14947
R:Gander, E.S.; Holmstroem, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross
Plant Mol. Biol. 16, 437-448, 1991
A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin fr
A:Reference number: S14946; MUID:91370890
A:Accession: S14947
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <GAN>
A:Cross-references: EMBL:X54491; NID:g17712; PIDN:CAA8363.1; PID:g17713
C:Genetics:
A:Introns: 67/3
C:Superfamily: wheat alpha-amylase inhibitor

Query Match 31.8%; Score 113.5; DB 2; Length 154;
Best Local Similarity 28.6%; Pred. No. 0.00011;
Matches 18; Conservative 21; Mismatches 21; Indels 3; Gaps 1;

QY 4 ESQQLQCCNOKVQVRDECCQCAIKYIAEDQIQGQLHGESERVAQRAGEIVSSCGV-- 61
DB 83 EEPHLDCCCEQLERMDCEKRCERGLRMMLRQRREMELOGEQMRLMKRAENLISRCNLS 142
QY 62 -RC 63
DB 143 QRC 145

RESULT 7
NMNKL
2S albumin 1 precursor - Arabidopsis thaliana
N:Alternate names: seed storage protein AT251
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 31-Mar-1992 #sequence, revision 31-Mar-1992 #text, change 03-Dec-1999
R:Kreibers, E.; Herdies, L.; de Clercq, A.; Seurinck, J.; Leemans, J.; Van Damme, J.; Se
Plant Physiol. 87, 859-866, 1988
A:Title: Determination of the processing sites of an Arabidopsis 2S albumin and characte
A:Reference number: JA0161
A:Accession: JA0161
A:Molecule type: DNA
A:Residues: 1-164 <KRE>
A:Cross-references: GB:M22032; NID:g166609; PIDN:AAA32743.1; PID:g166614
A:Accession: PS0282
A:Molecule type: protein
A:Residues: 38-73;84-162 <KR2>

R:Conceicao, A.D.S.; Kreibers, E.
submitted to the EMBL Data Library, July 1993
A:Description: Tentative title: a coryledon regulatory region is responsible for the

A:Reference number: S34674
A:Accession: S34676
A:Molecule type: DNA
A:Residues: 1-164 <CON>
A:Cross-references: EMBL:Z24745; NID:g395203; PIDN:CAA80870.1; PID:g395204
R:Bevan, M.; Van Der Schuren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.;
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15484
A:Accession: T06044

A:Molecule type: DNA
A:Residues: 1-164 <BEV>
A:Cross-references: EMBL:AL035680; GSPDB:GN00062; ATSP:T24A18.90
A:Experimental source: cultivar Columbia; BAC clone T24A18
C:Genetics:
A:Gene: T24A18.90
A:Map position: 4
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-164/Product: 2S albumin 1 proprotein #status predicted <AT2>
F:38-73/Product: 2S albumin 1 small chain #status experimental <SMC>
F:84-162/Product: 2S albumin 1 large chain #status experimental <LAC>

Query Match 30.8%; Score 110; DB 1; Length 164;
Best Local Similarity 34.4%; Pred. No. 0.00026;
Matches 21; Conservative 12; Mismatches 28; Indels 0; Gaps 0;

QY 1 QOESQQLQCCNOKVQVRDECCQCAIKYIAEDQIQGQLHGESERVAQRAGEIVSSCG 60
DB 87 QOEDQQLFQCCNELROBEPDVCPTLKQAKAVRLQGHQHPQVQRKTYQAKHLPNVDC 146
QY 61 V 61
DB 147 I 147

RESULT 8
T09252
seed storage protein EMB25 - white spruce
C:Species: Picea glauca (white spruce)
C:Date: 11-Jun-1999 #sequence, revision 11-Jun-1999 #text, change 11-Jun-1999
C:Accession: T09252
R:Dong, J.Z.; Dunstan, D.I.
submitted to the EMBL data library, June 1996
A:Description: Gene expression during somatic embryogenesis.
A:Reference number: Z16588
A:Accession: T09252
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-165 <DON>
A:Cross-references: EMBL:L47745; NID:g1350513; PID:g1350514
C:Genetics:
A:Gene: EMB25

Query Match 30.1%; Score 107.5; DB 2; Length 165;
Best Local Similarity 26.6%; Pred. No. 0.00047;
Matches 21; Conservative 20; Mismatches 21; Indels 17; Gaps 1;

QY 1 QOESQQLQCCNOKVQVRDECCQCAIKYIAED-----QIQGQLHGE 43
DB 66 EEPDQPSERCEELGRKSPCQRCQATQRTLEDFVMSDSQDGAFLNQRARRGRGGRGM 125
QY 44 ESEYVAQRAGEIVSSCGVR 62
DB 126 EEEYVRAAEELPMTCNVR 144

RESULT 9

R:Krebers, E.; Herdies, L.; de Clercq, A.; Seurinck, J.; Leemans, J.; Van Damme, J.; Se
Plant Physiol. 87, 859-866, 1988
A:Title: Determination of the processing sites of an Arabidopsis 2S albumin and character
A:Reference number: JAO161
A:Accession: JAO163
A:Molecule type: DNA
A:Residues: 1-164 <KRE>
A:Cross-references: GB:M22035; NID:g166611; PIDN:AAA32745.1; PID:g166616
R:Concilio, A.D.S.; Krebers, E.
Submitted to the EMBL Data Library, July 1993
A:Description: Tentative title: a cotyledon regulatory region is responsible for the dif
A:Reference number: S34674
A:Accession: S34674
A:Molecule type: DNA
A:Residues: 1-164 <CON>
A:Cross-references: EMBL:Z24744; NID:g395200; PIDN:CAA8068.1; PID:g395201
R:Tagata, A.; Kano, M.; Kawakami, M.; Ohki, Y.
Submitted to JIPID, December 1995
A:Description: Two dimensional electrophoresis of plant proteins and standardization of
A:Reference number: PN0173
A:Accession: PN0174
A:Molecule type: protein
A:Residues: 81-94 <TSU>
A:Experimental source: seeds
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voelt, M.; Robben, J.; Volckaert, G.; Ba
Submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15484
A:Accession: T06046
A:Molecule type: DNA
A:Residues: 1-164 <BEV>
A:Cross-references: EMBL:AL035680; GSPDB:GN00062; ATSP:T24A18.110
A:Experimental source: cultivar Columbia; BAC clone T24A18
C:Genetics:
A:Gene: T24A18.110
A:Map position: 4
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-164/Product: 2S albumin 3 protprotein #status predicted <AT7>
F:38-72/Product: 2S albumin 3 small chain #status predicted <SMC>
F:82-162/Product: 2S albumin 3 large chain #status predicted <LAC>

Query Match 28.3%; Score 101; DB 1; Length 164;
Best Local Similarity 36.6%; Pred. No. 0.0022;
Matches 26; Conservative 11; Mismatches 28; Indels 6; Gaps 3;
QY 2 QOESQLOQCCNQVRYDECCCEAIKYIAEDQIQGQLHGESESVYVQAGEIYSSCG 60
Db 86 QOQYQLQCCNENLRQEPVVCPTLQKAAAVASLQCG-HGPFQSRKRYQSAKYLPIICK 144
QY 61 VR---CMKQT 67
Db 145 IQQVGECPFGT 155

RESULT 14
B23617
conglutin delta-2 large chain - narrow-leaved blue lupine
C:Species: Lupinus angustifolius (narrow-leaved blue lupine)
C>Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 11-Jan-2000
C:Accession: B23617
R:Fillley, G.G.; Inglis, A.S.
FEBS Lett. 195, 235-241, 1986
A:Title: Amino acid sequence of conglutin delta, a sulfur-rich seed protein of Lupinus a
A:Reference number: A91358
A:Accession: B23617
A:Molecule type: protein
A:Residues: 1-80 <LII>
C:Superfamily: soybean 2S albumin

Query Match 27.5%; Score 98; DB 2; Length 80;

Best Local Similarity 32.9%; Pred. No. 0.0024;
Matches 23; Conservative 16; Mismatches 23; Indels 8; Gaps 4;
QY 1 QOESQLOQCCNQVRYDECCCEAIKYIAEDQIQGQLHGESESVYVQAGEIYSSCG 57
Db 7 ESESEEDQCCCEQINELNSORCOCRAEQIYESOSE--QCEGRQEQ--QLEGELEKLP 62
QY 58 -SCGVRCMRQ 66
Db 63 RICGFGPLRR 72

RESULT 15
T08010
2S seed storage protein precursor - Douglas fir
C:Species: Pseudotsuga menziesii (Douglas fir)
C>Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 11-May-2000
C:Accession: T08010
R:Chahal, M.; Misra, S.
Planta 206, 138-145, 1998
A:Title: Structure and expression of embryogenesis-specific cDNAs encoding 2S seed st
A:Reference number: Z16286; MUID:98381325
A:Accession: T08010
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-162 <CHA>
A:Cross-references: EMBL:AF029970; NID:g2988477; PIDN:AAC26998.1; PID:g2988478
C:Genetics:
A:Gene: 2S1
C:Keywords: seed; storage protein

Query Match 26.8%; Score 95.5; DB 2; Length 162;
Best Local Similarity 25.6%; Pred. No. 0.0081;
Matches 21; Conservative 21; Mismatches 21; Indels 19; Gaps 4;
QY 1 QOESQLOQCCNQVRYDECCCEAIKYI-----ABDQIQ--GQLHG--ESEE 46
Db 66 ERREDSSESNCNELERMSPCRCRPAIQVLDQASAFWDSDALNQRRGREGGRREG 125
QY 47 RVAQRAGEIYSSCGV----RC 63
Db 126 EMERAAVLPDTCNVQESPRRC 147

Search completed: September 9, 2002, 12:43:30
Job time: 222 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:39:48 ; Search time 45.25 Seconds
(without alignments)
78.570 Million cell updates/sec

Title: US-09-913-351-2

Perfect score: 186

Sequence: 1 ESKGERESSSQCRQEVQRKDLSCRCRYLRQSSSR 37

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186	100.0	258	1 RZCS	2S seed storage pr
2	73.5	40.6	141	2 T10257	2S albumin precurs
3	74	39.8	323	2 S38887	2S albumin - commo
4	70.5	37.9	154	2 S14947	2S albumin - Brazil
5	63.5	34.1	28	2 A25802	2S seed storage pr
6	62	33.3	749	2 S50095	splicing regulator
7	62	33.3	798	2 T15336	hypothetical prote
8	61.3	33.1	146	2 S14946	2S seed storage pr
9	59	31.7	295	2 S01062	2S seed storage pr
10	58.5	31.5	32	2 B59346	seed storage prote
11	58	31.2	148	2 S46514	putridolone-b pre
12	57	30.6	153	2 A33090	conglutin delta pr
13	57	30.6	699	2 E84565	hypothetical prote
14	56	30.1	37	2 A23617	conglutin delta-2
15	55	29.6	241	2 T27917	hypothetical prote
16	55	29.6	648	2 S04832	glutennin high mole
17	55	29.6	660	2 A24266	glutennin high mole
18	55	29.6	705	2 S18733	glutennin high mole
19	55	29.6	4815	2 JN0689	glutennin, high-mol
20	54	29.0	417	2 T47616	hypothetical prote
21	54	29.0	1377	2 T51447	transcription regu
22	53.5	28.8	100	2 G88637	protein F53H1.4 [i
23	53.5	28.8	104	2 S48180	mabninin IV - Yun
24	53.5	28.8	104	2 S48178	mabninin III - Yun
25	53.5	28.8	848	1 EEMTHW	glutennin, high mol
26	53.5	28.8	838	1 S02262	glutennin high mole
27	53	28.5	215	1 A47118	periplasmic ribonu
28	53	28.5	301	1 E71558	probable fad synth
29	53	28.5	374	2 T33328	hypothetical prote

30	53	28.5	518	2 D84977	glutamate--cystein
31	53	28.5	685	2 B82606	conjugal transfer
32	53	28.5	815	2 B30843	glutennin high mole
33	53	28.5	830	2 S15720	glutennin high mole
34	53	28.5	3844	2 T18402	asparagine/asparta
35	52	28.0	722	2 T02421	hypothetical prote
36	52	28.0	2061	2 T13751	transcription fact
37	51.5	27.7	104	2 S48176	mabninin I-1 - Yun
38	51.5	27.7	161	2 J04966	high-molecular-wel
39	51.5	27.7	165	2 T09252	seed storage prote
40	51.5	27.7	418	2 H83126	probable secretion
41	51.5	27.7	655	2 T28885	hypothetical prote
42	51	27.4	60	2 S29179	high-molecular-wel
43	51	27.4	63	2 S29177	high-molecular-wel
44	51	27.4	187	2 B83467	hypothetical prote
45	51	27.4	302	2 T13457	hypothetical prote

ALIGNMENTS

RESULT 1
RZCS
2S seed storage protein precursor - castor bean
N:Alternate names: 2S albumin precursor
C:Species: Ricinus communis (castor bean)
C:Date: 14-Nov-1983 #sequence, revision 08-Feb-1996 #text-change 18-Jun-1999
C:Accession: S11499; S11500; S11501; S11502; S27221; A01328; S27222
R:Irwin, S.D.; Keen, J.N.; Flindlay, J.B.C.; Lord, J.M.
Mol. Gen. Genet. 222, 400-408, 1990
A:Title: The authors translated the codon CTC for residue 14 as Phe, CCA for residue 7
A:Reference number: S11500; MUID:91109729
A:Accession: S11500
A:Molecule type: mRNA
A:Residues: 1-13, 'E', '15-73, 'T', '75-258 <IR2>
A:Experimental source: clone 14g4
A:Accession: S11501
A:Molecule type: mRNA
A:Residues: 'M', '4', 'LS', '7-13, 'F', '15-21 <IRF>
A:Experimental source: clone 10a12
A:Accession: S27221
A:Molecule type: protein
A:Residues: 'X', '37, 'X', '39-45, 'X', '158-161, 'X', '163-174, 'X' <IRA>
R:Shacter, F.S.; Li, S.S.U.
J. Biol. Chem. 257, 14753-14759, 1982
A:Title: Amino acid sequence of small and large subunits of seed storage protein from
A:Reference number: A92357; MUID:83082772
A:Accession: A01328
A:Molecule type: protein
A:Residues: 157-190; 194-221, 'Q', '223-225, 230-233, 'N', '235-254, 'Q', '256-258 <SHA>
A:Note: 230-Ser was also found
A:Note: there is considerable similarity between residues 181-231 of this protein and
R:Odani, S.; Koide, T.; Ono, T.; Ohnishi, K.
Biochem. J. 213, 543-545, 1983
A:Title: Structural relationships between barley (Hordeum vulgare) trypsin inhibitor a
A:Reference number: A90322; MUID:83308577
A:Contents: annotation
A:Note: this protein is homologous with trypsin inhibitor from barley
C:Complex: consists of two chains linked by two disulfide bonds involving Cys-162 and
C:Superfamily: 2S seed storage protein precursor

C:Keywords: pyroglutamic acid; seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-35/Domain: propeptide #status predicted <PRO>
F:36-72/Product: probable 2S seed storage protein small chain 2 #status experimental <SC
F:87-156/Product: probable 2S seed storage protein large 2 #status predicted <LCH
F:157-190/Product: 2S seed storage protein small chain #status experimental <SML>
F:194-258/Product: 2S seed storage protein large chain #status experimental <LNG>
F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 100.0%; Score 186; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 2,6e-16;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ESKGEREGSSQOCROEVQRKDLSSCERYLRQSSRR 37
DB 36 ESKGEREGSSQOCROEVQRKDLSSCERYLRQSSRR 72

RESULT 2
T10257
2S albumin precursor - cucurbit
N:Alternate names: prepro2S albumin
C:Species: Cucurbita sp. (cucurbit)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10257; S19323
R:Hara-Nishimura, I.; Takeuchi, Y.; Inoue, K.; Nishimura, M.
Plant J. 4, 793-800, 1993
A:Title: Vesicle transport and processing of the precursor to 2S albumin in pumpkin.
A:Reference number: Z17000; MUID:94100993
A:Accession: T10257
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA; protein
A:Residues: 1-141 <HAR1>
A:Cross-references: EMBL:D16560; NID:q459404; PIDN:BA03993.1; PID:q459405
A:Experimental source: seed, storage deposition stage; cotyledon
A>Note: soluble seed protein
R:Hara-Nishimura, I.; Inoue, K.; Nishimura, M.
FBS Lett. 294, 89-93, 1991
A:Title: A unique vacuolar processing enzyme responsible for conversion of several propt
A:Reference number: S19323; MUID:92077151
A:Accession: S19323
A:Status: preliminary
A:Molecule type: mRNA
C:Keywords: seed; storage protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-35/Domain: propeptide #status experimental <PRO>
F:36-141/Product: 2S albumin #status experimental <MAT>

Query Match 40.6%; Score 75.5; DB 2; Length 141;
Best Local Similarity 47.1%; Pred. No. 0.017;
Matches 16; Conservative 7; Mismatches 10; Indels 1; Gaps 1;

OY 1 ESKGEREGSSQOCROEVQRKDLSSCERYLRQSS 34
DB 31 ESKGEREGSSQOCROEVQRKDLSSCERYLRQSS 63

RESULT 3
S38887
2S albumin - common sunflower (fragment)
C:Species: Helianthus annuus (common sunflower)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-May-2000
C:Accession: S38887
R:Thyges, P.J.E.; Millichip, M.; Stobart, A.K.; Griffiths, W.T.; Napier, J.A.; Shewry, P.
submitted to the EMBL Data Library, November 1993
A:Reference number: S38887
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-523 <THO>

A:Cross-references: EMBL:X76101; NID:q429181; PID:q429182
C:Superfamily: gliadin

Query Match 39.8%; Score 74; DB 2; Length 323;
Best Local Similarity 48.5%; Pred. No. 0.056;
Matches 16; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

OY 6 REGSSQOCROEVQRKDLSSCERYLRQ--SSSR 36
DB 171 REGSSQOCROEVQRKDLSSCERYLRQSSSR 203

RESULT 4
S14947
2S albumin - Brazil nut
C:Species: Bertholletia excelsa (Brazil nut)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C:Accession: S14947
R:Gander, E.S.; Holmstroem, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gr
Plant Mol. Biol. 16, 437-448, 1991
A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin
A:Reference number: S14946; MUID:91370890
A:Accession: S14947
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <GAN>
A:Cross-references: EMBL:X54491; NID:g17712; PIDN:CAA38363.1; PID:g17713
A:Genetics:
A:Introns: 67/3
C:Superfamily: wheat alpha-amylase inhibitor

Query Match 37.9%; Score 70.5; DB 2; Length 154;
Best Local Similarity 42.4%; Pred. No. 0.079;
Matches 14; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

OY 1 ESKGEREGSSQOCROEVQRKDLSSCERYLRQ 32
DB 34 ESKGEREGSSQOCROEVQRKDLSSCERYLRQ 66

RESULT 5
A25802
2S seed storage protein small chain - Brazil nut
C:Species: Bertholletia excelsa (Brazil nut)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 19-May-2000
C:Accession: A25802
R:Ampe, C.; Van Damme, J.; de Castro, L.A.B.; Sampaio, M.J.A.M.; Van Montagu, M.; Van
Eur. J. Biochem. 159, 597-604, 1986
A:Title: The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil
A:Reference number: A91173; MUID:87004679
A:Accession: A25802
A:Molecule type: protein
A:Residues: 1-28 <AMP>
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 34.1%; Score 63.5; DB 2; Length 28;
Best Local Similarity 54.5%; Pred. No. 0.13;
Matches 12; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 12 OOCROEVQRKDLSSCERYLRQ 32
DB 2 OOCROEVQRKDLSSCERYLRQ 23

RESULT 6
S50095
splicing regulator Ceswap - Caenorhabditis elegans
C:Species: Caenorhabditis elegans

C>Date: 01-Aug-1995 #sequence_revision 24-Jul-1997 #text_change 17-Mar-1999
C:Accession: S53765; S50095
R:Spikes, P.A.; Kramer, J.; Bingham, P.M.; van Doren, K.
Nucleic Acids Res. 22, 4510-4519, 1994
A:Title: SMAP pre-mRNA splicing regulators are a novel, ancient protein family sharing a
A:Reference number: S50095; MUID:95061415
A:Accession: S53765
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-749 <SP1>
A:Cross-references: EMBL:U06932; NID:g459659; PID:g459660
A:Accession: S50095
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-551 <SP2>
A:Cross-references: EMBL:U06933
C:Genetics:
A:Introns: 56/2; 104/1; 434/3; 513/3
A>Note: this list of introns may be incomplete

Query Match 33.3%; Score 62; DB 2; Length 749;
Best Local Similarity 46.2%; Pred. No. 4.1;
Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 EREGSSOOCROEVOKRDKDSCERTL 30
DB 71 KRNGSPSEOCPTAEEDWCCEERTL 96

RESULT 7
T15336
Hypothetical protein B0336.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15336
R:Taich, A.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid B0336.
A:Reference number: Z18331
A:Accession: T15336
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-798 <TAI>
A:Cross-references: EMBL:U33305; NID:g912752; PID:g912761; PIDN:AA046835.1; CESP:B0336.9
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:B0336.9
A:Introns: 17/3; 60/2; 80/2; 128/1; 458/3; 537/3

Query Match 33.3%; Score 62; DB 2; Length 798;
Best Local Similarity 46.2%; Pred. No. 4.4;
Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 EREGSSOOCROEVOKRDKDSCERTL 30
DB 95 KRNGSPSEOCPTAEEDWCCEERTL 120

RESULT 8
S14946
2S seed storage protein large chain - Brazil nut
N:Alternate names: albumin 2S precursor
C:Species: Bertholletia excelsa (Brazil nut)
C>Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 21-Jul-2000
C:Accession: S14946; S14479; S06252; S21640; B25802
R:Gander, E.S.; Holmstrom, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross
Plant Mol. Biol. 16, 437-448, 1991
A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin fr
A:Reference number: S14946; MUID:91370890
A:Accession: S14946
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-146 <GAB>
A:Cross-references: EMBL:X54490; NID:g17710; PIDN:CAA38362.1; PID:g17711
A>Note: the authors translated the codon CTT for residue 13 as Val and GTC for residu
R:Bassener, R.; Schlesier, B.
submitted to the EMBL Data Library, December 1990
A:Reference number: S14479
A:Accession: S14479
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <BAS>
A:Cross-references: EMBL:X57027; NID:g17714; PID:g17715
R:Altenbach, S.B.; Pearson, K.W.; Leung, F.W.; Sun, S.S.M.
Plant Mol. Biol. 8, 239-250, 1987
A:Title: Cloning and sequence analysis of a cDNA encoding a Brazil nut protein except
A:Reference number: S06252
A:Accession: S06252
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-146 <ALT>
A:Experimental source: clone pHS-3
A>Note: part of this sequence, including the amino end of the large chain, was confir
R:Bassener, R.; Schlesier, B.
submitted to the EMBL Data Library, December 1990
A:Reference number: S21640
A:Accession: S21640
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <BA2>
A:Cross-references: EMBL:X57028; NID:g17716; PID:g17717
R:Ampe, C.; Van Damme, J.; de Castro, L.A.B.; Sampato, M.J.A.M.; Van Montagu, M.; Van
Eur. J. Biochem. 159, 597-604, 1986
A:Title: The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil
A:Reference number: A91173; MUID:87004679
A:Accession: B25802
A:Molecule type: protein
A:Residues: 70-90, 'E', 92-121, 'W', 123-125, 'L', 127-142 <AMP>
C:Genetics:
A:Introns: 60/3
C:Superfamily: wheat alpha-amylase inhibitor
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-46/Domain: propeptide #status predicted <PRO>
F:47-69/Product: seed storage protein small chain #status predicted <SCH>
F:70-146/Product: seed storage protein large chain #status experimental <LCH>

Query Match 33.1%; Score 61.5; DB 2; Length 146;
Best Local Similarity 40.7%; Pred. No. 1;
Matches 11; Conservative 8; Mismatches 7; Indels 1; Gaps 1;

OY 7 EGGSSOOCROEVOKRDKDLSCEERTLRQ 32
DB 33 EEEENOECREOMQROOMLSHCYMRQ 59

RESULT 9
S01062
2S seed storage protein precursor (clone Hag5) - common sunflower
N:Alternate names: 2S albumin storage protein
C:Species: Helianthus annuus (common sunflower)
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 21-Jul-2000
C:Accession: S01062
R:Allen, R.D.; Cohen, E.A.; Vonder Haar, R.A.; Adams, C.A.; Ma, D.P.; Nessler, C.L.;
Mol. Gen. Genet. 210, 211-218, 1987
A:Title: Sequence and expression of a gene encoding an albumin storage protein in sun
A:Reference number: S01062; MUID:88142538
A:Accession: S01062
A:Molecule type: DNA
A:Residues: 1-295 <ALL>
A:Cross-references: EMBL:X06410; NID:g18799; PIDN:CAA29699.1; PID:g18800
A>Note: part of this sequence, including the amino end of the mature protein, was con
C:Genetics:
A:Introns: 192/2

C; Superfamily: soybean 2S albumin

Query Match	30.1%;	Score 56;	DB 2;	Length 37;
Best Local Similarity	37.5%;	Pred. NO. 1.5;		
Matches	9;	Conservative	8;	Mismatches 7;
				Indels 0;
				Gaps 0;

```
QY 9 SSSQCCRQEVQRKDLSSCERYLRQ 32
    ||| |:::|:| ||:|
DB 3 SSEQCKRQLQOVNLRHCENHIDQ 26
```

RESULT 15
T27917

C:Accession: T27917
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Species: *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Protein: hypothetical protein ZK563.5 - *Caenorhabditis elegans*

R;le, T.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid ZK563.
A;Reference number: 730130

A;Reference number: Z20439
A;Accession: T27917

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA

A;Residues: 1-241 <LET>
A;Cross-references: EMB

A/Cross-References: EMBL:U40001; PDB:5AAB1D1.1; CSD:ZK303.3
C/Genetics:
A/Cat: OREGON:PRC3 F

A;Gene: CESP:ZK563.5
A;Introns: 41/3; 78/3; 130/1; 188/3; 203/3

C:Superfamily: Caenorhabditis elegans hypothetical protein ZK563.5

Query Match 20.68: Score 55: DB 2: Length 241:

Query Match	29.68;	Score 55;	DB 2;	Length 241;
Best Local Similarity	29.38;	Pred. No. 11;		
Matches 12; Conservative	7;	Mismatches 12;	Indels 10;	Gaps 1;

```

OY      5 EREGSSSQOCROE-----VQRKDLSCERYLR0SSS 35
      : || : || : || : || : || : || : || : || : ||
DB      77 DTGRIKQAOCCYEASAYRFLVRCVQKNISKCNFEFRKMS 117

```

Search completed: September 9, 2002, 12:43:30
Job time: 222 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:42:39 : Search time 94.6 Seconds
(without alignments)
79.842 Million cell updates/sec

Title: US-09-913-351-3
Perfect score: 357
Sequence: 1 QQQESQQLQCCNQVKQVND.....AQRAGEIVSSCGVRCMKQTR 68

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq_032802:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	100.0	68	21	AA15422
2	121	33.9	65	21	AA15423
3	110	30.8	164	10	AA191892
4	110	30.8	164	10	AA191892
5	107.5	30.1	158	18	AAW23588
6	106.5	29.8	158	18	AAW23586
7	98	27.5	171	15	AA193579
8	98	27.5	184	21	AA19926
9	96	26.9	184	12	AA14193
10	96	26.9	184	15	AA154981
11	96	26.9	184	16	AA176694

12	96	26.9	184	18	AA15415	Napin encoded by c
13	96	26.9	184	22	AA130526	Brassica campestris
14	96	26.9	184	22	AA185931	B. campestris lamb
15	96	26.9	189	9	AA181140	Sequence encoded b
16	96	26.3	142	21	AA154980	Napin protein from
17	94	26.3	142	21	AA184356	Amino acid sequenc
18	91.5	25.6	166	22	AA133600	Modified Ara h 2 a
19	91.5	25.6	166	22	AA105035	Modified anaphylac
20	91.5	25.6	167	22	AA104710	Modified anaphylac
21	91	25.5	184	19	AA130670	Napin seed storage
22	90.5	25.4	155	18	AA123418	Chimeric Soybean a
23	90.5	25.4	158	20	AA123420	Ara h 2 protein fr
24	87.5	24.5	156	20	AA140973	Peanut allergen Ar
25	87.5	24.5	157	18	AA124164	Peanut allergen Ar
26	87.5	24.5	157	18	AA124153	Peanut allergen Ar
27	87.5	24.5	157	20	AA115245	Peanut allergen, A
28	87.5	24.5	157	22	AA104707	Anaphylactic antiq
29	87.5	24.5	166	20	AA140968	Recombinant Ara h
30	87.5	24.5	207	22	AA182383	Peanut allergen Ar
31	86.5	24.2	155	18	AA110569	Mablinin II prepro
32	86.5	24.2	155	18	AA123587	Mablinin II prepro
33	85.5	23.9	93	12	AA141111	Brazil nut 12 kd s
34	85.5	23.9	110	15	AA162762	Synthetic 12 kd pr
35	85.5	23.9	111	19	AA186264	Wild-type Brazil n
36	85.5	23.9	111	19	AA186269	Modified Brazil nu
37	85.5	23.9	140	10	AA191891	Part of the sequen
38	84.5	23.7	111	19	AA186271	Modified Brazil nu
39	84.5	23.7	111	19	AA186267	Modified Brazil nu
40	84.5	23.7	111	19	AA186268	Modified Brazil nu
41	84.5	23.7	111	19	AA186270	Modified Brazil nu
42	84.5	23.7	148	21	AA158049	Puroindoline B pro
43	84.5	23.7	148	21	AA158049	Serine substituted
44	84.5	23.7	149	19	AA186272	Modified Brazil nu
45	84.5	23.7	149	19	AA186273	Modified Brazil nu

ALIGNMENTS

RESULT 1	
AA15422	AA15422 standard; peptide; 68 AA.
ID	AA15422
XX	AA15422;
AC	17-JAN-2001 (first entry)
XX	
DT	Castor bean 2S albumin storage protein amino acids 87-154.
XX	
DE	
XX	
KW	Antitumour; antilactaric; immunostimulatory; glycoconjugate; mannose;
KW	polysaccharide; mannose; galactose; castor bean; immune response; human;
KW	2S albumin storage protein; animal; tumour necrosis factor; mononuclear;
KW	phagocyte; granulocyte; macrophage; tumour; immunogen; Candida utilis.
OS	
XX	Ricinus communis.
PN	
XX	WO200050087-A1.
PD	
XX	31-AUG-2000.
PF	
XX	21-OCT-1999; 99MO-ES00338.
PR	
XX	26-FEB-1999; 99ES-0000408.
PA	
XX	(INFA-) IND FARM CANTABRIA SA.
PI	Brieva Delgado A; Garcia Villarrubia V; Guerrero Gomez-Pamo A;
PI	Piyel Rantieri JP; Gimenez Gallejo G; Matji Tuduri JA;
XX	
DR	WPI: 2000-558369/51.
XX	
PT	New glycoconjugate, useful for treating immunological disorders,
	comprises polysaccharide from Candida utilis and polypeptide from

```

PT Ricinus communis -
XX
XX
PS Claim 3; Page 30; 38pp; Spanish.
XX
CC The invention relates to a novel glycoconjugate (A) comprising a
CC 50-250 kDa polysaccharide (PS), containing one phosphate group per
CC 5-25 monosaccharide units, and a polypeptide (PP) having a particular
CC consensus sequence. In PS, at least 40% of the monosaccharide residues
CC are mannose and the remainder are glucose and/or galactose. The main
CC chain comprises 1-6 links, with less than 60% 1,2-branching. The PP has
CC the consensus sequence:
CC ZACZBD(QERK)2(Z')(LIVM)ZCCC(Z')(QEH)(LV)Z6CZCZ2ZdGZCZ(VILM)ZfCZg
CC where Z = any amino acid (aa); Z' = hydrophobic aa; Z'' = hydrophilic aa;
CC a = 3-48; b = 9-13; c = 15-39; d = 13-56; e = 15-26; f = 1-8; g = 1-11;
CC the parentheses indicate the preferred sequence. This sequence
CC corresponds to amino acids 87-154 of the Ricinus communis (castor bean)
CC 2S albumin storage protein which is used to generate the glycoconjugate
CC of the invention, with a polysaccharide from Candida utilis. (A) modifies
CC the immune response in humans and animals, especially it reduces the
CC amount of tumour necrosis factor (TNF) produced and stimulates the
CC mononuclear-phagocyte system and expands the granulocyte-macrophage
CC compartment. (A) are used for treating disorders of the immune system,
CC e.g. infections and tumours. (A) do not interfere with metabolic
CC processes in the liver and are active when given orally. PP significantly
CC increases the antigenic response to weakly immunogenic PS, including
CC induction of a T-dependent response.
XX
SQ Sequence 68 AA;
XX
Query Match 100.0%; Score 357; DB 21; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.6e-36;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 QOOSQOIQCCNCKVQKVRDECCCAIKYIAEDQIQOGOLHGESESRVQAGRIYSSCG 60
Db 1 qqqesqgqgqcnvkvqyrdcccaalkylaedqiqgqqlngeservaqagelvsqg 60
XX
QY 61 VRCKRQTR 68
Db 61 vrcmrqr 68
XX
RESULT 2
AAB15423
ID AAB15423 standard; peptide; 65 AA.
XX
AC AAB15423;
XX
DT 17-JAN-2001 (first entry)
XX
DE Castor bean 2S albumin storage protein amino acids 194-258.
XX
KW Antitumour; antimicrobial; immunostimulatory; glycoconjugate; mannose;
KW polysaccharide; mannose; galactose; castor bean; immune response; human;
KW 2S albumin storage protein; animal; tumour necrosis factor; mononuclear;
KW phagocyte; granulocyte; macrophage; tumour; immunogen; Candida utilis.
XX
OS Ricinus communis.
XX
PN W0200050087-A1.
XX
PD 31-AUG-2000.
XX
PF 21-OCT-1999; 99WO-ES00338.
XX
PR 26-FEB-1999; 99ES-0000408.
XX
PA (INFA-) IND FARM CANTABRIA SA.
XX
PI Brieva Delgado A, Garcia Villarrubia V, Guerrero Gomez-Pamo A;
PI Pivel Ranieri JF, Gimenez Gallego G, Matji Tuduri JA;
XX

```

```

DR WPI; 2000-558369/51.
XX
XX
PT New glycoconjugate, useful for treating immunological disorders.
PT comprises polysaccharide from Candida utilis and polypeptide from
PT Ricinus communis -
XX
XX
PS Claim 3; Page 30; 38pp; Spanish.
XX
CC The invention relates to a novel glycoconjugate (A) comprising a
CC 50-250 kDa polysaccharide (PS), containing one phosphate group per
CC 5-25 monosaccharide units, and a polypeptide (PP) having a particular
CC consensus sequence. In PS, at least 40% of the monosaccharide residues
CC are mannose and the remainder are glucose and/or galactose. The main
CC chain comprises 1-6 links, with less than 60% 1,2-branching. The PP has
CC the consensus sequence:
CC ZACZBD(QERK)2(Z')(LIVM)ZCCC(Z')(QEH)(LV)Z6CZCZ2ZdGZCZ(VILM)ZfCZg
CC where Z = any amino acid (aa); Z' = hydrophobic aa; Z'' = hydrophilic aa;
CC a = 3-48; b = 9-13; c = 15-39; d = 13-56; e = 15-26; f = 1-8; g = 1-11;
CC the parentheses indicate the preferred sequence. This sequence
CC corresponds to amino acids 194-258 of the Ricinus communis (castor bean)
CC 2S albumin storage protein which is used to generate the glycoconjugate
CC of the invention, with a polysaccharide from Candida utilis. (A) modifies
CC the immune response in humans and animals, especially it reduces the
CC amount of tumour necrosis factor (TNF) produced and stimulates the
CC mononuclear-phagocyte system and expands the granulocyte-macrophage
CC compartment. (A) are used for treating disorders of the immune system,
CC e.g. infections and tumours. (A) do not interfere with metabolic
CC processes in the liver and are active when given orally. PP significantly
CC increases the antigenic response to weakly immunogenic PS, including
CC induction of a T-dependent response.
XX
SQ Sequence 65 AA;
XX
Query Match 33.9%; Score 121; DB 21; Length 65;
Best Local Similarity 34.5%; Pred. No. 1.8e-07;
Matches 20; Conservative 15; Mismatches 23; Indels 0; Gaps 0;
XX
QY 4 ESQOIQCCNCKVQKVRDECCCAIKYIAEDQIQOGOLHGESESRVQAGRIYSSCGV 61
Db 1 qerslrgcchlkqmgqrcqglrqaiegsgqgqgqgqfcaftaanlpsmcgv 58
XX
RESULT 3
AAP91892
ID AAP91892 standard; protein; 164 AA.
XX
AC AAP91892;
XX
DT 29-APR-1990 (first entry)
XX
DE Amino acid sequence of 1kb fragment of the Arabidopsis thaliana
DE 2S-albumin gene.
XX
KW 2S-albumin; Arabidopsis thaliana; storage protein gene;
KW heterologous polypeptide.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FH Peptide 1..37
FH FT /note="signal peptide"
FH FT 38..73
FH FT /note="mature small subunit"
FH FT 74..83
FH FT /note="processing site"
FH FT 84..157
FH FT /note="mature large subunit"
XX
PN W08903887-A.
XX
XX
PD 05-MAY-1989.
XX

```


DR		WP1: 1989-159382/22.
DR	N-PSDB; AAN90116.	
XX		Transgenic plants contg. protein of high nutritional value -
PT	obtd. by inserting gene for 2S albumin modified for enrichment	
PT	of specific aminoacid(s)	
XX		
PS	Example; Fig 4; 31pp; English.	
CC		
CC	Prodn. of transgenic plants with increased nutritional value is claimed.	
CC	The method comprises cultivating plants, obtd. over 1 or more generations	
CC	from regenerated plant cells (or seeds) which include a nucleic acid	
CC	encoding a modified form of a natural 2S albumin storage protein, under	
CC	control of a promoter. The modified proteins are expressed at high	
CC	levels, without any alteration to correct expression, processing and	
CC	transport.	
XX		
SQ	Sequence 164 AA;	
QY		
Db		
Qy	1 QQQESQQLQCCNQKYVDECCGCAIRFYIAEDQIQGCQLHGESESRVAQRAGETIVSSCG 60 : : : : 87 qqqeqqlfqgcneirgeepdcvcpklkqaakavrlqghqgmgyrkikyqtakhlpncvd 146	
Db	147 I 147	
RESULT 5		
ID	AAMW23588	
AAW23588	AAMW23588 standard; Protein: 158 AA.	
XX		
AC	AAMW23588;	
XX		
DT	30-SEP-1997 (first entry)	
XX		
DE	Mablinin MBLIIT from Capparis masalakai.	
XX		
KW	MBL, sweetener; transgenic plant; recombinant protein; food;	
KW	beverage; animal feed; chewing gum; dental hygiene product;	
KW	pharmaceutical.	
XX		
OS	Capparis masalakai.	
XX		
FH	Key	
FT	Peptide	
FT	/label= Signal	
FT	Protein	
FT	/label= MBLIIT	
FT	Region	
FT	/label= A_chain	
FT	Region	
FT	/note= "14 amino acids evidently lost when mature dimer is formed"	
FT	Region	
FT	/label= B_chain	
XX		
PN	WC9700945-A1.	
XX		
PD	09-JAN-1997.	
XX		
PF	21-JUN-1996;	
XX	96WO-US10669.	
PR	23-JUN-1995;	
XX	95US-0000480.	
XX	(UYHA-) UNIV HAWAII.	
XI	Chen H, Hu Z, Sun SSM, Xiong L;	

XX WPI; 1997-087372/08.
XX Mabinlin nucleotide sequences - for prodn. of transgenic plants or
XX recombinant proteins, useful as sweeteners, partic. in foods
XX
XX PS Disclosure; Fig 3; 38pp; English.
XX
XX The present sequence represents the mabinlin MBLII. The encoding cDNA
XX can modulate the expression of the gene encoding MBL in a cell or
XX tissue. Mabinlin comes from the seeds of the Capparis masakal plant,
XX and is limited by its native availability, and so methods for producing
XX recombinant mabinlin protein are being produced using chimeric genes
XX such as MBLII (see AAT60776). The products provide recombinant materials
XX for the production of the MBL protein in practical amounts, and for the
XX production of transgenic plants containing inherently sweet edible parts
XX by virtue of production of MBL in situ. The MBL protein can be used to
XX sweeten foods, beverages, animal feeds, chewing gum, dental hygiene
XX products and pharmaceuticals.
XX
XX Sequence 158 AA;
SQ
Query Match 30.1%; Score 107.5; DB 18; Length 158;
Best Local Similarity 32.8%; Pred. No. 2.3e-05;
Matches 22; Conservative 13; Mismatches 27; Indels 5; Gaps 2;
OY 2 Q0ESQLOQCCNOVQVDECCCEAIKRYIADQIQGOLHG-EESERYAQRAGETIVSSCG 60
DB 87 qqrpalrlcncqrlrqnkvpcvpylrqahqqllygqgieprvrrqlfiraarnlpnck 146
OY 61 V----RC 63
DB 147 lpavgrc 153
RESULT 6
AAW23586
ID AAW23586 standard; Protein; 158 AA.
XX
XX AAW23586;
XX
XX 30-SEP-1997 (first entry)
XX
XX Mabinlin MBLI from Capparis masakal.
XX
XX MBL; sweetener; transgenic plant; recombinant protein; food;
XX beverage; animal feed; chewing gum; dental hygiene product;
XX pharmaceutical.
XX
XX Capparis masakal.
XX
XX Key Location/Qualifiers
XX FH Peptide 1..39
XX FT /label= Signal
XX FT 40..158
XX FT /label= MBLI
XX FT 40..70
XX FT /label= A_chain
XX FT 71..86
XX FT /note= "14 amino acids evidently lost when mature
XX dimer is formed"
XX FT Region 87..158
XX FT /label= B_chain
XX
XX W09700945-A1.
XX
XX 09-JAN-1997.
XX
XX 21-JUN-1996; 96WO-US10669.
XX
XX 23-JUN-1995; 95US-0000480.
XX

PA (UYHA-) UNIV HAWAII.
XX
XX Chen H, Hu Z, Sun SSM, Xiong L;
XX WPI; 1997-087372/08.
XX
XX DR N-PSDB; AAT60775.
XX
XX Mabinlin nucleotide sequences - for prodn. of transgenic plants or
XX recombinant proteins, useful as sweeteners, partic. in foods
XX
XX PS Disclosure; Fig 3; 38pp; English.
XX
XX
XX The present sequence represents the mabinlin MBLI. The encoding cDNA can
XX modulate the expression of the gene encoding MBL in a cell or tissue.
XX Mabinlin comes from the seeds of the Capparis masakal plant, and
XX is limited by its native availability, and so methods for producing
XX recombinant mabinlin protein are being produced using chimeric genes
XX such as MBLII (see AAT60776). The products provide recombinant materials
XX for the production of the MBL protein in practical amounts, and for the
XX production of transgenic plants containing inherently sweet edible parts
XX by virtue of production of MBL in situ. The MBL protein can be used to
XX sweeten foods, beverages, animal feeds, chewing gum, dental hygiene
XX products and pharmaceuticals.
XX
XX Sequence 158 AA;
SQ
Query Match 29.8%; Score 106.5; DB 18; Length 158;
Best Local Similarity 34.3%; Pred. No. 3e-05;
Matches 23; Conservative 10; Mismatches 29; Indels 5; Gaps 2;
OY 2 Q0ESQLOQCCNOVQVDECCCEAIKRYIADQIQGOLHGEESERYAQRAG-ETIVSSCG 60
DB 87 qqrpalrlcncqrlrqnkvpcvpylrqahqqllygqgieprvrrqlfiraarnlpnck 146
OY 61 V----RC 63
DB 147 lpavgrc 153
RESULT 7
AAR53579
ID AAR53579 standard; Protein; 171 AA.
XX
XX AAR53579;
XX
XX 25-OCT-1994 (first entry)
XX
XX Synthetic 2S seed storage protein.
XX
XX Nutrition; non-conserved DNA; assembly; crude; degenerate;
XX oligonucleotides.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FH Peptide 39..74
XX FT /note= "small subunit"
XX FT 85..170
XX FT /note= "large subunit"
XX FT 1..38
XX FT /note= "signal peptide"
XX FT 75..84
XX FT /note= "linker excised as protein folds"
XX FT MISC-difference 171
XX FT /note= "excised at protein maturity"
XX
XX W09410315-A.
XX
XX 11-MAY-1994.
XX
XX 22-OCT-1993; 93WO-US10090.
XX

```

PR 23-OCT-1992; 92US-0965664.
XX
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Ballo B;
XX
XX WPI: 1994-167470/20.
XX N-PSDB: AAQ63710.
XX
XX Improving the content of an amino acid in a seed storage protein
PT to enhance nutritional value - by replacing non-conserved DNA
PT with DNA encoding the amino acid
XX
XX
XX Disclosure; Fig 1; 33pp: English.
XX
XX The sequence is that of a 2S seed storage protein with increased
CC lysine content. The protein is synthesised from a set of partial
CC DNA sequences capable of being assembled in a complementary over-
CC lapping relationship to provide the complete DNA. The protein is
CC translated as the straight chain shown and undergoes post translational
CC modification to give the mature protein which has two subunits joined by
CC a di-sulphide bond.
XX
XX Sequence 171 AA;
XX
XX
XX Query Match 27.5%; Score 98; DB 15; Length 171;
XX Best Local Similarity 30.3%; Pred. No. 0.00037;
XX Matches 20; Conservative 16; Mismatches 24; Indels 6; Gaps 2;
XX
XX 2 Q0ESQI0QCCNQVQKVRDECCQCE---AIKYLAEQI0QCG0LHGEES--ERVAQRAGEI 55
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 89 qqrplp1qkccceqlkqmgsgcvcplkqgaskavkqgeggqggqgqkqgmvrklyktakhl 148
XX
XX 56 VSSCGV 61
XX | :
XX 149 pkrvcdl 154
XX
XX
XX RESULT 8
XX AA49926
XX ID AA49926 standard; Protein: 184 AA.
XX
XX AC AA49926;
XX
XX
XX 31-JAN-2000 (first entry)
XX
XX Brassica campestris lambda-CGM1-2 napin protein sequence.
XX
XX Brassica: regulation; development; identification; screening;
XX light promoter; seed promoter; fruit promoter; regulatory region;
XX napin; soybean; rapeseed; cotton; safflower; sunflower; expression;
XX phenotype; modification.
XX
XX Brassica campestris.
XX
XX US5981839-A.
XX
XX 09-NOV-1999.
XX
XX
XX 07-MAR-1997; 97US-0812665.
XX
XX
XX 17-JAN-1985; 85US-0692605.
XX 23-JAN-1988; 88US-0147781.
XX 29-APR-1988; 88US-0188361.
XX 02-NOV-1988; 88US-0267865.
XX 09-JUL-1990; 90US-0550804.
XX 10-AUG-1993; 93US-0105852.
XX 07-JUN-1995; 95US-0484941.
XX 31-JUL-1986; 86US-0891529.
XX 26-MAY-1987; 87US-0054369.
XX 28-JUL-1987; 87US-0078538.
XX 15-MAR-1988; 88US-0168190.
XX

```

```

PR    21-MAY-1990;      90US-0526123.
PR    14-SEP-1990;      90US-0582241.
PR    08-AUG-1991;       91US-0742834.
XX
PA    (CALJ ) CALGENE LLC.
XX
PI    Kridl JC, Knauf VC;
XX
DR    WPI; 2000-012275/01.
XX    N-PSDB; AA235691.
XX
PT    New genetic constructs useful for transforming plants with a DNA
sequence of interest -
XX
XX    Example 3; Fig 2; 65pp; English.
XX
PS    The present invention describes genetic constructs (I) for transforming
CC    plants with a DNA sequence of interest, comprising a promoter from a
CC    napin, Eay or acyl carrier protein gene, a DNA sequence of interest and
CC    flanking T-DNA. The constructs can be used for the expression of
CC    heterologous genes in plants. They can be used for modifying the
CC    phenotype of plants. The DNA sequence of interest may encode an enzyme
CC    or may comprise an antisense sequence. The constructs can be used in
CC    plants such as soybean, rapeseed, cotton, safflower or sunflower. The
CC    present sequence represents the Brassica campestris lambda-CGN1.2 napin
CC    protein.
CC
SQ    Sequence   184 AA;

Query Match          27.5%; Score 98; DB 21; Length 184;
Best Local Similarity 35.5%; Pred. No. 0.0004;
Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;

QY     1 OOOEEEOO-----LQQCCNOVKQVRDSCCCEAIK---YIAEDOI--OOG-OLHGEESSE---- 46
      ||| || | | | | | | | | | | : | | | | | | | | : | | | | | | | | :
Db     92 qqqqpqqpplllqqccnnheepclvecptlkgaakavqqlrqqgqqgmqqmqqyvis 151
      | : | | : : | : |
QY     47 RVACRAGEIVSSCGV R 62
      | : | | : : | : |
Db     153 rltqtatnpractnlr 167

RESULT 9
AAR14193
ID AAR14193 standard; Protein; 184 AA.
XX
AC AAR14193;
XX
DT 11-DEC-1991 (first entry)
XX
DE Napin 1-2 gene product.
XX
KW Desaturase; expression cassette; B. campestris.
XX
PN WO9113972-A.
XX
PD 19-SEP-1991.
XX
PF 14-MAR-1991; 91WO-U001746.
XX
PR 14-NOV-1990; 90US-0615784.
PR 16-MAR-1990; 90US-0494106.
PR 13-AUG-1990; 90US-0567373.
XX
PA (CALG-) CALGENE INC.
XX
PI Thompson G, Knauf V;
XX
DR WPI; 1991-295627/40.
XX    N-PSDB; AAQ13969.
XX
PT DNA encoding a plant desaturase - used for modifying the satd.
```

PT fatty acid compsn. of plant cells and plant seeds
XX
PS Disclosure; Fig 10; 128pp; English.
XX
CC An expression construct contg. a plant desaturase suitable for plant
CC transformation is produced. 5' upstream and 3' downstream sequences
CC obtainable from the B. campestris napin gene here are used.
CC See also AA013963-69.
XX
SQ Sequence 184 AA;

Query Match 26.9%; Score 96; DB 12; Length 184;
Best Local Similarity 35.5%; Pred. No. 0.0007;
Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;

OY 1 QOQESQ0---LQCCNQVQVRDECCCEAIK---YIAEDQI--QOG-QLHGESE-----46
||| || |||||::: | | | : || || |::: |
DB 92 qqqgpgqrppllqgcnelhgeepclvcgplkqaskavkqgllrqgqgqgmqqvls 151
OY 47 RVAORAGEIVSSGCVR 62
| : | | : : | : |
DB 152 rlyqtathlpracnir 167

RESULT 10
AAR54981
ID AAR54981 standard; Protein; 184 AA.
XX
AC AAR54981;
XX
DT 07-DEC-1994 (first entry)
XX
DE Napin protein from Brassica campestris.
XX
KW enhance seed oil production; acyl carrier protein; ACP-I.
XX
OS Brassica campestris.
XX
PN US5315001-A.
XX
PD 24-MAY-1994.
XX
PF 31-JUL-1986; 86US-0891529.
XX
PR 31-JUL-1986; 86US-0891529.
PR 28-JUL-1987; 87US-0078824.
PR 15-NOV-1989; 89US-0437764.
PR 28-JAN-1992; 92US-0826696.
XX
PA (CALJ) CALGENE INC.
XX
PI Knauf VC, Kridl JC;
XX
DR WPI; 1994-166646/20.
XX
DR N-PSDB; AAQ65428.
XX
PT DNA encoding plant acyl carrier protein - used to enhance prodn.
PT of seed oil and to modulate the fatty acid compsn. of the oil
XX
PS Disclosure; Fig 4E-4G; 28pp; English.
XX
CC AAR54981 shows a Brassica napin protein, generated from lambda CGN1-2
CC clone. The napin gene promoter is regulated during seed maturation.
CC This can be used to control the expression of an ACP in seed to
CC enhance the production of seed oil and modulate the fatty acid
CC composition. The DNA can also be used to prepare probes.
XX
SQ Sequence 184 AA;

Query Match 26.9%; Score 96; DB 15; Length 184;
Best Local Similarity 35.5%; Pred. No. 0.0007;

Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;

OY 1 QOQESQ0---LQCCNQVQVRDECCCEAIK---YIAEDQI--QOG-QLHGESE-----46
||| || |||||::: | | | : || || |::: |
DB 92 qqqgpgqrppllqgcnelhgeepclvcgplkqaskavkqgllrqgqgqgmqqvls 151
OY 47 RVAORAGEIVSSGCVR 62
| : | | : : | : |
DB 152 rlyqtathlpracnir 167

RESULT 11
AAR76694
ID AAR76694 standard; Protein; 184 AA.
XX
AC AAR76694;
XX
DT 04-NOV-1995 (first entry)
XX
DE B. campestris napin.
XX
KW Napin; promoter; seed-specific gene transcription; oilseed;
KM transgenic plant.
XX
OS Brassica campestris.
XX
PN US5420034-A.
XX
PD 30-MAY-1995.
XX
PF 31-JUL-1986; 86US-0891529.
XX
PR 09-JUL-1990; 90US-0550804.
PR 31-JUL-1986; 86US-0891529.
PR 28-JUL-1987; 87US-0078838.
PR 25-JAN-1988; 88US-0147781.
PR 08-AUG-1991; 91US-0742834.
XX
PA (CALJ) CALGENE INC.
XX
PI Knauf VC, Kridl JC;
XX
DR WPI; 1995-206250/27.
DR N-PSDB; AAQ94586.
XX
PT DNA providing seed specific gene transcription - includes a
PT transcription initiator from a gene preferentially expressed in
PT seed cells, e.g. the napin gene, also related expression cassettes
PT and Brassica plants.
XX
PS Disclosure; Fig.2A-J; 40pp; English.
XX
CC Clone lambda CGN1-2 was isolated by screening a partial genomic
CC library of B. campestris DNA for napin-encoding clones. The entire
CC CDS as well as extensive 5' and 3' sequences of the napin gene were
CC sequenced (AAQ94586). An expression cassette was constructed from
CC the 3' - and 5'-ends of the lambda CGN1-2 napin gene.
XX
SQ Sequence 184 AA;

Query Match 26.9%; Score 96; DB 16; Length 184;
Best Local Similarity 35.5%; Pred. No. 0.0007;
Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;

OY 1 QOQESQ0---LQCCNQVQVRDECCCEAIK---YIAEDQI--QOG-QLHGESE-----46
||| || |||||::: | | | : || || |::: |
DB 92 qqqgpgqrppllqgcnelhgeepclvcgplkqaskavkqgllrqgqgqgmqqvls 151
OY 47 RVAORAGEIVSSGCVR 62
| : | | : : | : |
DB 152 rlyqtathlpracnir 167

```

XX RESULT 12
XX AAW15415
XX ID AAW15415 standard; Protein, 184 AA.
XX AC
XX AAW15415;
XX DT 05-JUN-1997 (first entry)
XX DE Napiin encoded by clone gamma-ACGN1-2.
XX KW Brassica; fatty acid; rat thioesterase; acyl ACP;
XX KW phospholipid acyl desaturase.
XX OS Brassica campestris.
XX PN
XX US5608152-A.
XX PD 04-MAR-1997.
XX PF 31-JUL-1986; 86US-0891529.
XX PE 09-JUL-1990; 90US-0550804.
XX PR 31-JUL-1986; 86US-0891529.
XX PR 28-JUL-1987; 87US-0078538.
XX PR 25-JAN-1988; 88US-0147781.
XX PR 08-AUG-1991; 91US-0742834.
XX PR 30-MAY-1995; 95US-0453924.
XX PA (CALJ ) CALGENE INC.
XX PI Knauf VC, Kridl JC;
XX DR WPI; 1997-164560/15.
XX DR N-PSDB; AAT65475.
XX PT Transgenic Brassica plant and seeds - contain DNA construct for seed
XX PT specific expression of heterologous proteins
XX PS
XX PS Example 2; Column 23-28; 48pp; English.
XX CC The sequences given in AAW15415-17 represent proteins encoded by DNA
XX CC fragments which were used in the construction of the DNA construct
XX CC of the invention. The DNA construct comprises, in the 5' to 3'
XX CC direction of transcription:
XX CC (a) a transcriptional initiation region from a gene that encodes a
XX CC product preferentially expressed in a plant seed cell as compared to
XX CC other plant cells;
XX CC (b) a DNA sequence of interest other than the native coding sequence
XX CC of the gene; and
XX CC (c) a transcriptional termination region;
XX CC where the gene is a napiin gene, an acyl carrier protein gene or an EA9
XX CC gene. This construct may be used in the production of a Brassica plant
XX CC which may be modified such that heterologous proteins are produced in
XX CC the seeds, such as fatty acids, rat thioesterase, acyl ACP or
XX CC phospholipid acyl desaturase.
XX Sequence 184 AA;
XX

```

	Query Match	26.9%	Score 96;	DB 18;	Length 184;	
	Best Local Similarity	35.5%;	Pred. No. 0.0007;			
	Matches 27; Conservative	12;	Mismatches 23;	Indels 14;	Gaps 5;	
OY	1	QQQEQ-----LQCCNKKVRCDCGCEAIK---YIADQI--QG-QLHGEESE----	46			
			:::			:
			:::			:
Db	92	qqqgppqprippllqqccmelhgpeclvcptlkgsakvkgqlrqqgqgmqqmqqvls	151			
OY	47	RVAQRAGEIVSSCGVR	62			
	:	:	:	:	:	:
Db	152	riygtathtipracnrl	167			

RESULT	13	
AAW30526		
ID	AAW30526	standard; Protein; 184 AA.
XX		
XX	AAW30526;	
AC		
XX		
XX	26-OCT-1998	(first entry)
DT		
XX		
DE	Brassica campestris napin seed storage protein.	
XX		
KW	Transgenic plant; tissue-specific gene expression; napin gene;	
XX	promoter; seed storage protein.	
OS	Brassica campestris.	
XX		
PN	US5753475-A.	
PD	19-MAY-1998.	
XX		
PF	10-AUG-1993;	93US-0105852.
XX		
PR	10-AUG-1993;	93US-0105852.
PR	17-JAN-1985;	85US-0692805.
PR	31-JUL-1986;	86US-0891529.
PR	26-MAY-1987;	87US-0054369.
PR	28-JUL-1987;	87US-0078538.
PR	25-JAN-1988;	88US-0147781.
PR	15-MAR-1988;	88US-0168190.
PR	29-APR-1988;	88US-0188361.
PR	02-NOV-1988;	88US-0267685.
PR	21-MAY-1990;	90US-0526123.
PR	14-SEP-1990;	90US-0582241.
PR	08-AUG-1991;	91US-0742834.
XX		
PA	(CALJ) CALGENE INC.	
XX		
PL	Houck CM;	
XX		
DR	WPI: 1998-311403/27.	
DR	N-PSDB; AAV36968.	
XX		
PT	Transformation of plants - with regulatory sequence containing	
PT	constructs for tissue specific expression of genes	
XX		
PS	Example 3; Fig 2E-G; 67pp; English.	
XX		
CC	This is the napin seed storage protein of Brassica campestris,	
CC	encoded by genomic DNA in clone lambda CGN1-2 (see AAV36968). The	
CC	promoter region of the napin gene can be used in a novel method for	
CC	producing a plant with a regulatable phenotype. In this method,	
CC	regulatory regions from plant genes expressed during a particular	
CC	developmental stage or in a specific tissue are identified by cDNA	
CC	screening. The regulatory regions (and T-DNA and Ti or Ri plasmids)	
CC	are manipulated for use with foreign sequences for introduction	
CC	into plant cells to provide transformed plants having a phenotypic	
CC	property that can be modulated. The invention is exemplified with	
CC	light, seed and fruit-specific promoters. Also claimed are methods	
CC	for altering the phenotype of fruit tissue as distinct to other	
CC	plant tissue, modifying the genotype of a plant to impart a desired	
CC	characteristic to fruit, modifying transcription in fruit tissue,	
CC	and expressing a heterologous DNA sequence of interest specifically	
CC	in fruit tissue. The method provides transformation without gall	
CC	formation in plants which have historically not been Agrobacterium	
XX	hosts.	
XX		
Sequence	184 AA:	
30		

Query Match	26.9%	Score 96;	DB 19;	Length 184;
Best Local Similarity	35.5%	Pred. No. 0.0007;		
Matches 27; Conservative	12;	Mismatches 23;	Indels 14;	Gaps 5;

D6	87	QQRGALRLCCNQLRQVKKPCVCEVLNQAAHQQLYQGIQSPRQVRRLLFRAARLNPICK	146
QY	61	V----RC	63
	:		
D6	147	IPAVGR	153

RESULT 2
 US-08-670-186-4
 Sequence 4: Application US/08670186
 Patent No. 5859343
 GENERAL INFORMATION:
 APPLICANT: SUN, SAMUEL S.M.
 APPLICANT: XIONG, LIMEN
 APPLICANT: HU, ZHONG
 APPLICANT: CHEN, HANG
 TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500
 CITY: WASHINGTON
 STATE: DC
 COUNTRY: USA
 ZIP: 20006-1888
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/670,186
 FILING DATE: 21-JUN-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: MURASHIGE, KATE H.
 REGISTRATION NUMBER: 29,959
 REFERENCE/DOCKET NUMBER: 23461-20007.00
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 887-1500
 TELEFAX: (202) 822-0168
 TELEX: 90-4030 MRSNFORSMH
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 158 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-670-186-4

Query Match	29.88	Score 106.5	DB 2	Length 158
Best Local Similarity	34.38	Pred. No. 3.6e-06		
Matches 23	Conservative 10	Mismatches 29	Indels 5	Gaps 2

0y 2 QQESQQLQCCNQVQVRDECCCAITYIAEDYIQGGHGEESERVAQRG-EIYSSCG 60
|| : : ||| : || : : ||| : ||| : : :
Db 87 QQRGPAALRLCCNQLRVYNKPCPCPVLRLQAHQQLYQGGIEGRQVRLFRARNLPIICK 146

QY	61	V	---	RC	63
		:			
Db	147	IPAVGRC			153

RESULT 3
 US-08-453-924-3
 : Sequence 3, Application US/08453924
 : Patent No. 5608152
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Kridl, Jean C.
 : APPLICANT: Knauf, Vic C.
 :
 : TITLE OF INVENTION: Seed-Specific Transcriptional Regulation
 :
 : NUMBER OF SEQUENCES: 17

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Well, Gotshal & Manges
STREET: 2882 Sand Hill Road, Ste. 280
CITY: Menlo Park
STATE: CA
COUNTRY: U.S.A.
ZIP: 94025

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,924
FILING DATE: 30-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/742,834
FILING DATE: 08-AUGUST-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/550,804
FILING DATE: 09-JULY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/147,781
FILING DATE: 25-JANUARY-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/078,538
FILING DATE: 28-JULY-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/891,529
FILING DATE: 31-JULY-1986
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE-037/04US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6200
TELEX: (415) 854-3713
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 184 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-924-3

Query Match          26.9%: Score 96; DB 1; Length 184;
Best Local Similarity 35.5%; Pred. No. 9.le-05;
Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;

QY      1 OOOESQQ-----LQCCNNOVKQRDECCCEAIIK--YYAEQDI-QQG-QLHGESE-----46
        |||   |----|||:::|         |       :  ||   || :: |: 
Db      92 QQQGQGQPRLPILQCNCNELHGEPICVPTLKAGSKAVKKQIRQQGGQMGGQMQGVYS 151
        |||   |---:|::|         |       :  ||   || :: |: 

QY      47 RVAPRAGEIVSSCGVR 62
        |:-|-:-:|::|         |       :  ||   || :: |: 

Db     152 RLYQTATHLPACNR 167

RESULT      4
US-08-618-911-2
Sequence 2, Application US/08618911
Patent No. 5850016
GENERAL INFORMATION:
APPLICANT: Jung, Rudolf
APPLICANT: Hastings, Craig
APPLICANT: Coughlan, Sean
APPLICANT: Hu, David
TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN SEEDS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
```


ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,911
FILING DATE: Concurrently herewith
ATTORNEY/AGENT INFORMATION:
NAME: Simon, Soma
REGISTRATION NUMBER: 37,444
REFERENCE/DOCKET NUMBER: 365-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-618-911-2

Query Match 25.4%; Score 90.5; DB 2; Length 155;
Best Local Similarity 25.0%; Pred. No. 0.00036;
Matches 16; Conservative 25; Mismatches 16; Indels 7; Gaps 3;

QY 1 QQESQQLQCCNCKVQKQVDE-CQCEAIKYIAEDQIQGQLHGESESVAAQAGEIVSSC 59
Db 84 EEERGHMKCCSMELSKPICCKKALQKIMQNSE--QLEGKKQKQERE---LMLN 137

QY 60 GYRC 63
Db 138 AIRC 141

RESULT 5
US-08-618-911-6
Sequence 6, Application US/08618911
Patent No. 5850016
GENERAL INFORMATION:
APPLICANT: Jung, Rudolf
APPLICANT: Hastings, Craig
APPLICANT: Coughlan, Sean
APPLICANT: Hu, David
TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
STREET: 700 Capital Square, 400 Locust Street
CITY: Des Moines
STATE: Iowa
COUNTRY: USA
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,911
FILING DATE: Concurrently herewith
ATTORNEY/AGENT INFORMATION:
NAME: Simon, Soma
REGISTRATION NUMBER: 37,444

REFERENCE/DOCKET NUMBER: 365-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (515) 248-4800
TELEFAX: (515) 248-4844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-618-911-6

Query Match 25.4%; Score 90.5; DB 2; Length 158;
Best Local Similarity 25.4%; Pred. No. 0.00037;
Matches 15; Conservative 24; Mismatches 17; Indels 3; Gaps 2;

QY 2 QQESQQLQCCNCKVQKQVDE-CQCEAIKYIAEDQIQGQLHGESESVAAQAGEIVSSC 59
Db 88 EEERGHMKCCSMELSKPICCKKALQKIMQNSE--ELEEKKKMKEMLMNLAITMC 144

RESULT 6
US-08-670-186-2
Sequence 2, Application US/08670186
Patent No. 5859343
GENERAL INFORMATION:
APPLICANT: SUN, SAMUEL S.M.
APPLICANT: XIONG, LIWEN
APPLICANT: HU, ZHONG
APPLICANT: CHEN, HANG
TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,186
FILING DATE: 21-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MORASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 23461-20007.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSWSH
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-670-186-2

Query Match 24.2%; Score 86.5; DB 2; Length 155;
Best Local Similarity 27.9%; Pred. No. 0.0012;
Matches 17; Conservative 15; Mismatches 28; Indels 1; Gaps 1;

QY 2 QQESQQLQCCNCKVQKQVDE-CQCEAIKYIAEDQIQGQLHGESESVAAQAGEIVSSC 60
Db 83 QPRRPAIRGCCNQLRGVDRCCVCPVLQRAAQOYVLRGIIIGPQQLRRLLFPAARRLPICN 142

OY 61 V 61
DB 143 I 143

RESULT 7

US-08-618-911-4
; Sequence 4, Application US/08618911
; Patent No. 5850016
; GENERAL INFORMATION:
; APPLICANT: Jung, Rudolf
; APPLICANT: Hastings, Craig
; APPLICANT: Coughlan, Sean
; APPLICANT: Hu, David
; TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN
; NUMBER OF INVENTIONS: SEEDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,911
; FILING DATE: Concurrently herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Simon, Soma
; REGISTRATION NUMBER: 37,444
; REFERENCE/DOCKET NUMBER: 365-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-618-911-4

Query Match 23.4%; Score 83.5; DB 2; Length 158;

Best Local Similarity 23.7%; Pred. No. 0.0028;

Matches 14; Conservative 23; Mismatches 19; Indels 3; Gaps 2;

OY 2 Q0ESQ0LQCCNCKVQVRD-EC0CEAIKTYAEDQIQG0LHGESESRVQRAGEIVSSC 59

DB 88 EEEEGHMKCTEMSELSRSPKCKCKALOKIMENSE--ELEEKKKMKELINLATWC 144

RESULT 8

US-08-938-675A-2
; Sequence 2, Application US/08938675A
; Patent No. 6107287
; GENERAL INFORMATION:
; APPLICANT: de Lumen, Benito O.
; APPLICANT: Galvez, Alfredo F.
; TITLE OF INVENTION: Lunasin Peptides
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA

ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,675A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B98-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-938-675A-2

Query Match 23.4%; Score 83.5; DB 3; Length 158;

Best Local Similarity 23.7%; Pred. No. 0.0028;

Matches 14; Conservative 23; Mismatches 19; Indels 3; Gaps 2;

OY 2 Q0ESQ0LQCCNCKVQVRD-EC0CEAIKTYAEDQIQG0LHGESESRVQRAGEIVSSC 59

DB 88 EEEEGHMKCTEMSELSRSPKCKCKALOKIMENSE--ELEEKKKMKELINLATWC 144

RESULT 9

US-08-991-300-2
; Sequence 2, Application US/08991300
; Patent No. 5973225
; GENERAL INFORMATION:
; APPLICANT: D'VIDIO, RENATO
; APPLICANT: PORCEDDU, ENRICO
; APPLICANT: MERCHITELLI, CINZIA
; APPLICANT: CARDELLI, LUISA ERCOLI
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE
; TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,300
; FILING DATE: 16-DEC-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IT MI 96/A 002663
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2264-0201-0X
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 369 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-991-300-2

Query Match 20.7%; Score 74; DB 2; Length 369;
Best Local Similarity 32.8%; Pred. No. 0.13;
Matches 21; Conservative 12; Mismatches 21; Indels 10; Gaps 3;

OY 2 QOESQOLQCCNOVKYRDECCCEAIKYA-----EDQIQGQLHGESEVAQRAQ 53
Db 235 QOESQHWQCCQCOQLPQIPQOSRYEAIYVILIQEQOYVGSIQIQOQO--PQQLG 292
OY 54 EIVS 57
Db 293 QCVS 296

RESULT 10
US-08-625-209A-2
Sequence 2, Application US/08625209A
Patent No. 5756671
GENERAL INFORMATION:
APPLICANT: Gyuris, Jeno
APPLICANT: Lamphere, Lou
APPLICANT: Diarella, Giulio
TITLE OF INVENTION: Cd37 Cell-Cycle Regulatory Protein,
TITLE OF INVENTION: and Uses Related Thereto
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley, Hoag & Eliot
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/625,209A
FILING DATE: 01-APR-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MI048.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 373 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-625-209A-2

Query Match 20.6%; Score 73.5; DB 1; Length 373;
Best Local Similarity 27.8%; Pred. No. 0.15;
Matches 20; Conservative 15; Mismatches 28; Indels 9; Gaps 2;
OY 2 QOESQOLQCCNOVKYRDECCCEAIKYIABDQIQG-----QLHGESEVAQRAQRIV 56

Db 44 OKKEEILDRGCRCKRKRAVECO-----RKTKLEVAEGGKAELERLQASTAAAGGALG 99
OY 57 SSCGVRCMRQTR 68
Db 100 AEAGRCARRRR 111

RESULT 11
US-07-955-905A-23
Sequence 23, Application US/07955905A
Patent No. 5770433
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
TITLE OF INVENTION: PRECURSOR
NUMBER OF SEQUENCES: 28
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (PRO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/955,905A
FILING DATE: 21-JAN-1993
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 587 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Gossypium hirsutum
FEATURE:
NAME/KEY: Protein
LOCATION: 1..587
OTHER INFORMATION: /note= "Vicilin from G. hirsutum"
US-07-955-905A-23

Query Match 19.9%; Score 71; DB 1; Length 587;
Best Local Similarity 26.5%; Pred. No. 0.54;
Matches 18; Conservative 15; Mismatches 29; Indels 6; Gaps 1;

OY 1 QOESQOLQCCNOVKYRDECCCEAIKYI-----ABDQIQGQLHGESEVAQRAQ 54
Db 83 QRRYECCQOECROEEROPQCCORCIKRFEOQOQOOSOFQOECQOHQOEOERPERKOQ 142
OY 55 IVSSCGVR 62
Db 143 CVAECRRER 150

RESULT 12
US-08-531-927B-2
Sequence 2, Application US/08531927B
Patent No. 5840491
GENERAL INFORMATION:
APPLICANT: Kakizuka, Akira
TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
TITLE OF INVENTION: Disease Gene and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

1  COMPUTER: IBM PC compatible
2  OPERATING SYSTEM: PC-DOS/MS-DOS
3  SOFTWARE: PatentIn Release #1.0, Version #1.30
4  CURRENT APPLICATION DATA:
5  APPLICATION NUMBER: US/08/531,927B
6  FILING DATE: 21-SEP-1995
7  CLASSIFICATION: 435
8  PRIOR APPLICATION DATA:
9  APPLICATION NUMBER: JP H6-251600
10 FILING DATE: 21-SEP-1994
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Granahan, Patricia
13 REGISTRATION NUMBER: 32,227
14 REFERENCE/DOCKET NUMBER: ATH95-01A
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: 617-861-6240
17 TELEFAX: 617-861-9540
18 INFORMATION FOR SEQ ID NO: 2:
19 SEQUENCE CHARACTERISTICS:
20 LENGTH: 360 amino acids
21 TYPE: amino acid
22 TOPOLOGY: linear
23 MOLECULE TYPE: protein
24
25 US-08-531-927B-2

```

Query Match	18.5%;	Score 66;	DB 2;	Length 360;
Best Local Similarity	34.8%;	Pred. No. 1.3;		
Matches 23;	Conservative	8;	Mismatches 21;	Indels 14;
				Gaps 3;

OY 1 QQDSQQQLDCCNQKQVRFECCEALTYIAEDPIQGGLHGEES---ERVAORACIEIVS 57
|||: || | : | : | : | : | : | : | : |
Db 292 QQKQKQQQQQ---QQQQQQQQQQ-----QQQQQQRLSGSSHCPCERATSSGALGS 340

QY	58	SCGVR	63
Db	341	DLGKAC	346

RESULT 13
US-09-041-886-13
; Sequence 13, Application US/09041886

GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Rabilzadsh, Sharoz
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 7000

```

; STATE: California
; COUNTRY: United States
; ZIP: 92122

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/0041,886
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

```

; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
;

```

REFERENCE/DOCKET NUMBER: P-
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001

TELEPHONE: (619) 535-2001
TELEFAX: (619) 535-8949
INFORMATION FOR SEO ID NO: 13:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-13

```

Query Match	18.5%;	Score 66;	DB 4;	Length 360;
Best Local Similarity	34.8%;	Pred. No. 1.3;		
Matches	23;	Conservative	8;	Mismatches 21;
			Indels	14;
			Gaps	3;

QY 1 QQQEEOQLQCCNQVAVRDEQCCEATNYIAEDIQGGGLHGEES---ERAQAQRACELIVS 57
||| : ||| : | : : | : | : | : | : | : |
Db 292 QQQKQ000000---Q000000000-----Q00000RLSGSSHPCEPATSTGALGS 340

QY	58	SCGVR	63
Db	341	DLGKAC	346

RESULT 14
US-08-447-411-45
; Sequence 45, Application US/08447411

Patent NO. 5113243
GENERAL INFORMATION:
ADD TO ANTE. FIRST

APPLICANT: BREDEHORST, REINHARD
APPLICANT: VOGEL, CARL-WILHELM

```

; TITLE OF INVENTION:  DNA ENCODING COBRA C3, CVF1, AND CVF2
;
; NUMBER OF SEQUENCES:  81
;

```

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington
STATE: Virginia
COUNTRY: U S A

COMPUTER READ
ZIP: 22202
CONKRI: 0

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

;
; OPERATING SYSTEM: PC-DOS/MS
; SOFTWARE: PatentIn Release

```

```
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER:  US/08/447,411
;; FILING DATE:
```

FILING DATE: 435
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:

```

      FROM APPLICATION DATA.
      APPLICATION NUMBER:  US 08/043,747
      FILING DATE:  07-APR-1993

```

ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5773243man F.

```

;
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-101-0
;

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR THE... (5)

```

; INFORMATION FOR SEQ ID NO: 4
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1642 amino acids

```

LENGTH: 1042 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
; US-08-447-411-45

Query Match 18.28; Score 65; DB 1; Length 1642;

Best Local Similarity 41.4%; Pred. NO. 11;
Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY 10 QCCNQVRDECQCEAIKYIAEDQIQG 38

Db 710 ECCRYIKGVDEMORESELFILARDNEDG 738

RESULT 15

US-08-662-227-2
; Sequence 2, Application US/08662227
; Patent No. 5922320
; GENERAL INFORMATION:
; APPLICANT: VOGEL, CARL WILHELM
; APPLICANT: BREDEHORST, REINHORST
; APPLICANT: KOCK, MICHAEL
; APPLICANT: FRITZINGER, DAVID
; TITLE OF INVENTION: RECOMBINANT PROCVF
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/662,227
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-0107-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1642 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-662-227-2

Query Match 18.2%; Score 65; DB 2; Length 1642;
Best Local Similarity 41.4%; Pred. NO. 11;
Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

OY 10 OCCNOVKQVRDECCCAIKYIAEDQIQG 38
;||:| ||||| |: :| | :|
Db 710 ECCRYIKGVDEMORESELFILARDNEDG 738

Search completed: September 9, 2002, 12:44:13
Job time: 235 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:46:06 ; Search time 22.97 Seconds
(without alignments)
114.625 Million cell updates/sec

Title: US-09-913-351-3
Perfect score: 357
Sequence: 1 000ES0010QCCNQVKQVRD.....AQRAGETIVSCGVRCMRQTR 68

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357	100.0	258	1	2SS_RICCO
2	125	35.0	295	1	P01089 riclinus com
3	110	30.8	164	1	2SS5_HELAN
4	104.5	29.3	100	1	2SS1_ARATH
5	104.5	28.3	104	1	2SS4_CAPMA
6	103.5	28.0	104	1	2SS3_CAPMA
7	102.5	28.7	130	1	2SS1_CAPMA
8	101	28.3	164	1	ITRY_STINAR
9	98	27.5	80	1	2SS3_ARATH
10	91	25.5	178	1	CG2L_LUPAN
11	91	25.5	180	1	2SS2_BRANA
12	90.5	25.4	129	1	2SS3_BRANA
13	89	24.9	170	1	2SS2_ARATH
14	88	24.6	172	1	2SS1_ARATH
15	87	24.4	186	1	2SS1_PICGL
16	86.5	24.2	155	1	2SS2_BRANA
17	86	24.1	127	1	2SS2_CAPMA
18	85.5	23.9	146	1	ALH1_STINL
19	85.5	23.9	319	1	2SS_BEREX
20	84.5	23.7	148	1	GDAS_WHEAT
21	84.5	23.7	166	1	PU1B_WHEAT
22	84	23.5	178	1	2SS4_ARATH
23	83.5	23.4	110	1	2SSB_BRANA
24	83.5	23.4	158	1	2SS1_BRANA
25	83	23.2	223	1	2SS_SOYB
26	83	23.2	313	1	GLU2_MAIZE
27	82.5	23.1	289	1	GDAT_WHEAT
28	81.5	22.8	282	1	HOG3_HORVU
29	81.5	22.8	327	1	GDAS_WHEAT
30	79	22.1	304	1	GDAS_WHEAT
31	78.5	22.0	291	1	GOBB_WHEAT
32	76	21.3	220	1	GOBB_WHEAT
33	75.5	21.1	356	1	GLTA_WHEAT

34	75	21.0	307	1	GLTB_WHEAT	P10386 triticum ae
35	73	20.4	251	1	GDAS_WHEAT	P08079 triticum ae
36	72.5	20.3	297	1	GDAS_WHEAT	P04724 triticum ae
37	72.5	20.3	302	1	GDAS_WHEAT	P21292 triticum ae
38	72	20.2	244	1	GDAS_WHEAT	P04730 triticum ae
39	71.5	20.0	147	1	IAA_HORVU	P16969 hordeum vul
40	71.5	20.0	186	1	GL19_ORYZA	P29835 oryza sativ
41	71	19.9	305	1	HOG1_HORVU	P17990 hordeum vul
42	70.5	19.7	149	1	PRO7_ORYZA	P20698 oryza sativ
43	70	19.6	361	1	KOPF_SYNY3	P72830 synechocyst
44	69	19.3	143	1	IA16_WHEAT	P16159 triticum ae
45	69	19.3	291	1	GDAS_WHEAT	P04722 triticum ae

ALIGNMENTS

RESULT	ID	2SS_RICCO	STANDARD	PRT	258 AA.
AC	P01089	21-JUL-1986 (Rel. 01, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	25 albumin precursor (Allergen Ric c 1).				
OS	Ricinus communis (Castor bean).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid 1; Malpighiales; Euphorbiaceae; Ricinus.				
OX	NCBI_TaxID=3988;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Endosperm;				
RX	MEDLINE=91109729; PubMed=2274038;				
RA	Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;				
RT	"The Ricinus communis 2S albumin precursor: a single preproprotein				
RT	may be processed into two different heterodimeric storage proteins.";				
RL	Mol. Gen. Genet. 222:400-408(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Endosperm;				
RX	MEDLINE=91016940; PubMed=2216785;				
RA	Irwin S.D., Lord J.M.;				
RT	"Nucleotide sequence of a Ricinus communis 2S albumin precursor				
RT	gene.";				
RL	Nucleic Acids Res. 18:5890-5890(1990).				
RN	[3]				
RP	SEQUENCE OF 157-190 AND 194-258.				
RX	MEDLINE=83082772; PubMed=7174664;				
RA	Sharf F.S., Li S.S.-L.;				
RT	"Amino acid sequence of small and large subunits of seed storage				
RT	protein from Ricinus communis.";				
RL	J. Biol. Chem. 257:14753-14759(1982).				
RN	[4]				
RP	SIMILARITY TO PROTEINASE INHIBITORS.				
RX	MEDLINE=83308577; PubMed=6615448;				
RA	Odani S., Koide T., Ono T., Ohnishi K.;				
RT	"Structural relationship between barley (Hordeum vulgare) trypsin				
RT	inhibitor and castor-bean (Ricinus communis) storage protein.";				
RL	Biochem. J. 213:543-545(1983).				
CC	-1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.				
CC	-1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN				
CC	LINKED BY 2 DISULFIDE BONDS.				
CC	-1- PTM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE				
CC	CHAINS INVOLVE CYS-162 AND CYS-175.				
CC	-1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				

```
CC or send an email to license@lsb.sib.ch).
CC -----
DR EMBL: X54158; CAA38097.1; -.
DR PIR: A01328; RZCS.
DR PIR: S11499; S11499.
DR PIR: S11500; S11500.
DR PIR: S11501; S11501.
DR InterPro: IPR003612; AAI.
DR Pfam: PF00234; tryp_alpha_aml1_inh.
DR SMART: SM00499; AAI: 2.
DR SMART: SM00499; AAI: 2.
KW Seed storage protein; Signal; Allergen.
FT SIGNAL 1 21 PROBABLE.
FT PROPEP 22 156
FT CHAIN 157 190 2S ALBUMIN, SMALL CHAIN.
FT PROPEP 191 193
FT CHAIN 194 258 2S ALBUMIN, LARGE CHAIN.
FT MOD_RES 194 194 PYRROLIDONE CARBOXYLIC ACID.
FT CONFLICT 222 222 E -> Q (IN REF. 3).
FT CONFLICT 226 229 MISSING (IN REF. 3).
FT CONFLICT 234 234 D -> N (IN REF. 3).
FT CONFLICT 255 255 E -> Q (IN REF. 3).
FT SEQUENCE 258 AA; 29290 MW; 27874CFC50EA1072 CRC64;

Query Match
Best Local Similarity 100.0%; Score 357; DB 1; Length 258;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 00ESQQLQCCNOVKYRDECCCEAIKTYIAEDQIQGQLHGSESEVAORAGELVSSCG 60
DB 87 00ESQQLQCCNOVKYRDECCCEAIKTYIAEDQIQGQLHGSESEVAORAGELVSSCG 146
QY 61 VRCKRQTR 68
DB 147 VRCKRQTR 154

RESULT 2
2SS5_HELAN STANDARD; PRT; 295 AA.
ID 2SS5_HELAN
AC P15461;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 2S seed storage protein precursor (2S albumin storage protein).
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 162-173.
RC STRAIN=CV, GIANT GREY STRIPE;
RX MEDLINE=88142538; PubMed=2830455;
RA Allen R.D., Cohen E.A., Vonder Haar R.A., Adams C.A., Ma D.P.,
RA Nessler C.L., Thomas T.L.;
RT "Sequence and expression of a gene encoding an albumin storage
RT protein in sunflower."
RL M01. Genet. 210:211-218(1987).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- PWM: IT IS POSSIBLE THAT THE 38 KDA PRECURSOR IS CLEAVED INTO TWO
CC POLYPEPTIDES THAT ARE APPROXIMATELY THE SAME SIZE. THE MATURE
CC PROTEIN IS COMPOSED OF A SINGLE POLYPEPTIDE CONTAINING ONE OR MORE
CC INTRA-MOLECULAR DISULFIDE LINKAGES.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb.sib.ch/announce/
CC or send an email to license@lsb.sib.ch).
CC -----
DR EMBL: X06410; CAA29699.1; -.
DR PIR: S01062; S01062.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_aml1_inh.
DR Pfam: PF00234; tryp_alpha_aml1; 2.
DR SMART: SM00499; AAI: 2.
KW Seed storage protein; Albumin; Signal; Multigene family.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 161
FT CHAIN 162 295 2S SEED STORAGE PROTEIN.
FT SEQUENCE 295 AA; 34071 MW; 8958A106805142A1 CRC64;

Query Match
Best Local Similarity 31.1%; Score 125; DB 1; Length 295;
Matches 23; Conservative 22; Mismatches 19; Indels 10; Gaps 2;

QY 2 00ESQQLQCCNOVKYRDECCCEAIKTYIAEDQIQGQLHGSESEVAORAGELVSSCG 60
DB 74 RQEQQLQCCCELONTBGCCCEAVKQVFRRAQQVQGGGRLVPRRSGQRTQLKQK 133
QY 52 AGEIVSSGVRCMR 65
DB 134 AQLPNCVNLQSR 147

RESULT 3
2SS1_ARATH STANDARD; PRT; 164 AA.
ID 2SS1_ARATH
AC P15457;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S seed storage protein 1 precursor (2S albumin storage protein)
DE (NM002-2S albumin 1).
GN AT2S1 OR AT4G27140 OR T24A18.90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eusterois II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-73 AND 84-162.
RC STRAIN=CV, C24;
RA Krebbers E., Herdies L., de Clercq A., Seurinck J., Leemans J.,
RA van Damme J., Segura M., Gheysen G., van Montagu M.,
RA Vandekerckhove J.;
RT "Determination of the processing sites of an Arabidopsis 2S albumin
RT and characterization of the complete gene family."
RL Plant Physiol. 87:859-866(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, C24;
RA Conceicao A.D.S., Krebbers E.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Meyer K.F.X., Schueller C., Mambutt R., Murphy G., Volckaert G.,
RA Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weischelgartner M., de Simone V., Obermayer B., Maché R., Mueller M.,
RA Kreis M., Delseny M., Putigomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Gymonpretz B., Chuang T.-J., Vandebussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
```


RA Meltznerger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzner E., Brandt A., Peters S., van Staveren M., Dirse W.,
RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koelter P.,
RA Berneriser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buijssemaet C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., Mcclay K., Mayes R.,
RA Pettelt A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
RA Botkova D., Bloeker H., Scharte M., Grimm M., Loenert T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fatmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argitlou A., Vitale D., Liguori R., Pitravandi E.,
RA Massenot O., Ougley F., Clabaud G., Mendein A., Reiber R.,
RA Schnabl S., Hiller R., Schmidt W., Leclary A., Aubourg S.,
RA Cherdor F., Cooke R., Berger C., Montfort A., Casacubeta E.,
RA Glibons T., Weber N., Vandenbol M., Baquies M., Tereol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Haeremann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiigel L.,
RA Sekhon M., Murray J., Sheel P., Cordes M., Abu-Theilaid J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Gelsel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA Antonouli B., Zidanif M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shaker M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., Mccombie W.R.,
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999)).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC -1- LINKED BY DISULFIDE BONDS.
CC -1- MISCELLANEOUS: THIS IS THE MOST ABUNDANT ISOFORM OF 2S ALBUMIN IN
CC ARABIDOPSIS.
CC
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/e1sb-ch>).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M22032; AAA32743.1; -
DR EMBL: Z24745; CAA08070.1; -
DR EMBL: A13820; CAA01132.1; -
DR EMBL: AL035680; CAB59844.1; -
DR EMBL: AL161565; CAB79569.1; -
DR PIR: JAO161; NRMWU1.
DR PIR: S34676; S34676.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_1typ_aml_1nh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; 1typ_alpha_aml; 1.
DR PRINTS: PR00496; NAPIN.
DR ProDom: PD002498; Napin; 1.
DR SMART: SM004499; AAI; 1.
KW Seed storage protein; Albumin; Signal; Multigene family.
FT SIGNAL 1 21
FT PROPEP 22 37
FT CHAIN 38 73 2S SEED STORAGE PROTEIN 1 SMALL SUBUNIT.
FT PROPEP 74 83
FT CHAIN 84 162 2S SEED STORAGE PROTEIN 1 LARGE SUBUNIT.
FT PROPEP 163 164
SQ SEQUENCE 164 AA; 19014 MW; 2BF28CBA74D9832B CRC64;

```

Query Match 30.8%; Score 110; DB 1; Length 164;
Best Local Similarity 34.4%; Pred. No. 9,6e-05;
Matches 21; Conservative 12; Mismatches 28; Indels 0; Gaps 0;

QY 1 000ESQQLQCCNQYKQVDECCQCAIKYIADDOIQGQLHGESEERVAQRAGEIVSSCG 60
    ||| | ||||| : | : | : ||| : : | : |
Db 87 000EQQLPQCCCNELRQEPDVCVPTLQAKAVRLQGHQDMQYRKIVQYAKHLPNVSD 146
QY 61 V 61
Db 147 I 147

RESULT 4
2SS4_CAPMA STANDARD: PRT: 100 AA.
ID 2SS4_CAPMA AC P80353;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin IV, A and B chains (MAB IV) (Sweet protein).
OS Capparis maseikai (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Capparid.
OX NCBI_TaxID=13395;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RA MEDLINE=94333405; PubMed=8055976;
RX Nirasawa S., Nishino T., Katahira M., Desugi S., Hu Z., Kurihara Y.;
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.";
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
CC INDUCING ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR ProDom: PD002498; Napin.
DR SMART: SM00499; AAI; 1.
DR Seed storage protein; Albumin; Sweet-taste.
KW CHAIN
FT CHAIN 1 28 MABINLIN IV, A CHAIN.
FT NON_CONS 28 29 MABINLIN IV, B CHAIN.
FT CHAIN 29 100
FT DISULFID 4 49 BY SIMILARITY.
FT DISULFID 17 38 BY SIMILARITY.
FT DISULFID 39 87 BY SIMILARITY.
FT DISULFID 51 95 BY SIMILARITY.
SQ SEQUENCE 100 AA; 11928 MW; 7B09673FAB7793CA CMC64;

Query Match 29.3%; Score 104.5; DB 1; Length 100;
Best Local Similarity 31.3%; Pred. No. 0.00022;
Matches 21; Conservative 14; Mismatches 27; Indels 5; Gaps 2;

QY 2 000ESQQLQCCNQYKQVDECCQCAIKYIADDOIQGQLHGESEERVAQRAGEIVSSCG 60
    : | : ||||| : | : | : ||| : : | : |
Db 29 E0RGPALRCLCNOLKQVKKPCVCPVLRQAHQOLYQGIQEPQVRRFLFRARLNPICK 88
QY 61 V---RC 63
    : |
Db 89 IPAVGRC 95

```

2SS3_CAPMA	STANDARD:	PRT:	104 AA.
ID 2SS3_CAPMA			
AC P80352.			
DT 01-FEB-1995 (Rel. 31, Created)			
DT 01-FEB-1995 (Rel. 31, Last sequence update)			
DT 16-OCT-2001 (Rel. 40, Last annotation update)			
DE Mabinlin III, A and B chains (MAB III) (Sweet protein).			
OS Capparis masaiikai (Mabinlang)			
OC Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;			
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC eucotsids II; Brassicales; Brassicaceae; Capparis.			
OX NCBI_TaxID=13395;			
RN [1]			
RP SEQUENCE.			
RC TISSUE=Seed;			
RX MEDLINE=94333405; PubMed=8055976;			
RA Nirasawa S., Nishino T., Katalhira M., Uesugi S., Hu Z., Kurihara Y.,			
RZ "Structures of heat-stable and unstable homologues of the sweet			
RT protein mabinlin. The difference in the heat stability is due to			
RT replacement of a single amino acid residue."			
RL Eur. J. Biochem. 223:989-995(1994).			
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGEE PROTEIN HAVING SWEETNESS-			
CC INDUCING ACTIVITY.			
CC -1- SUBUNIT: HETERO-DIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY			
CC DISULFIDE BONDS.			
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGEE ALBUMINS FAMILY.			
CC Interpro: IPR003612; AAI.			
DR Interpro: IPR000617; Napln.			
DR Pfam: PF01631; Seedstore_2s; 1.			
DR PRINTS: PR00496; NABIN.			
DR Prodom: PD002498; Napln; 1.			
DR SMART: SM00499; AAI; 1.			
FT Seed storage protein; Albumin, Sweet-taste.			
FT CHAIN 1 32 MABINLIN III, A CHAIN.			
FT NON_CONS 32 33			
FT CHAIN 33 104 MABINLIN III, B CHAIN.			
FT DISULFID 4 53			
FT DISULFID 17 42			
FT DISULFID 43 91			
FT DISULFID 55 99			
SQ SEQUENCE 104 AA; 12284 MW; 102EE3F5E24AD3D0 CRC64;			
Query Match 29.3%; Score 104.5; DB 1; Length 104;			
Best Local Similarity 31.3%; Pred. No. 0.00023;			
Matches 21; Conservative 14; Mismatches 27; Indels 5; Gaps 2;			
OY 2 OQESQLOGCCNOVQVDECCCEAIKITYAEDQIOGGOLHG-EESERYAORAGETIVSSCG 60			
DB 33 EDRGALRRCNCNQLQVKNQKPCVCFVLRQAAHQOQLYGGQIEGFRVRRLLFRRAARNLPNICK 92			
QY 61 V---RC 63			
DB 93 IPAVGRC 99			
RESULT 6			
2SS1_CAPMA STANDARD: PRT: 104 AA.			
ID 2SS1_CAPMA			
AC P80351;			
DT 01-FEB-1995 (Rel. 31, Created)			
DT 01-FEB-1995 (Rel. 31, Last sequence update)			
DT 16-OCT-2001 (Rel. 40, Last annotation update)			
DE Mabinlin I-1, A and B chains (MAB I) (Sweet protein).			
OS Capparis masaiikai (Mabinlang).			
OC Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;			
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC eucotsids II; Brassicales; Brassicaceae; Capparis.			
OX NCBI_TaxID=13395;			
RN [1]			
RP SEQUENCE.			
RC TISSUE=Seed;			
RX MEDLINE=94333405; PubMed=8055976;			

```

RA Nirasawa S., Nishino T., Katakura M., Uesugi S., Hu Z., Kurihara Y.,
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue."
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: 2S SEED STORAGE PROTEIN HAVING SWEETNESS-INDUCING
CC ACTIVITY. THIS FORM IS NOT HEAT STABLE.
CC -1- SUBUNIT: HETERO DIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2s; 1.
DR PRINTS: PR00496; NAPIIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
KW Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 32 MABINLIN I, A CHAIN.
FT NON_CONS 32 33
FT CHAIN 33 104 MABINLIN I, B CHAIN.
FT DISULFD 4 53
FT DISULFD 17 42
FT DISULFD 43 91
FT DISULFD 55 99
SQ SEQUENCE 104 AA; 12284 MW; 156A4A4B50D5EE60 CRC64;

Query Match 29.0%; Score 103.5; DB 1; Length 104;
Best Local Similarity 32.8%; Pred. No. 0.00029;
Matches 22; Conservative 11; Mismatches 29; Indels 5; Gaps 2;

QY 2 QQESQQLQCCNQVQVQVRECCCEAKYIAEQIQGGLHGESESRVQNRAG-ETVSSCG 60
DB 33 EQRGALRCPCCNLRQVKNPCVCPVLRAAHQQLYQGGEGPRVQLLEARNLPNICK 92
QY 61 V---RC 63
DB 93 IPAVERC 99

RESULT 7
ITRY_SINAR
ID ITRY_SINAR STANDARD: PRT: 130 AA.
AC P38037;
DC 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin Inhibitor (TISA).
OS Sinapis arvensis (Charlock).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Sinapis.
OC NCBI_TaxID=29728;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=94350545; Pubmed=8070965;
RA Svendsen I.B., Nicolova D., Goshev I., Genov N.;
RT "Primary structure, spectroscopic and inhibitory properties of a two-
RT chain trypsin inhibitor from the seeds of charlock (Sinapis arvensis
RT L), a member of the napin protein family";
RL Int. J. Pept. Protein Res. 43:423-430(1994).
CC -1- FUNCTION: INHIBITS TRYPSIN WITH A KI OF 7 X 10(-6) M.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2s; 1.
DR PRINTS: PR00496; NAPIIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
KW Seed storage protein; Multigene family; Polymorphism.

```

```
FT CHAIN 1 39 SMALL CHAIN.
FT NON CONS 39 40
FT CHAIN 40 130 LARGE CHAIN.
FT VARIANT 32 32 R -> M.
FT VARIANT 53 57 MISSING (IN FORM II).
FT VARIANT 73 73 A -> S.
FT VARIANT 77 77 K -> R.
FT VARIANT 81 81 Q -> R.
FT VARIANT 87 87 H -> Q.
FT VARIANT 89 89 Q -> H.
FT VARIANT 91 91 E -> Q.
FT VARIANT 97 97 E -> M.
FT VARIANT 98 98 I -> V.
FT VARIANT 99 99 R -> S.
FT VARIANT 106 106 T -> K.
FT VARIANT 123 123 N -> Q.
FT VARIANT 124 124 K -> G.
FT VARIANT 126 126 M -> V.
SQ SEQUENCE 130 AA; 14682 MW; EC02EA826D180DF2 CRC64;

Query Match 28.7%; Score 102.5; DB 1; Length 130;
Best Local Similarity 35.2%; Pred. No. 0.00045;
Matches 25; Conservative 10; Mismatches 25; Indels 11; Gaps 3;

QY 2 QGESQDLOCCNQVNRDECCCEAKIYIA---EQOI-QQGQHGHESE-----RVNQ 50
DB 44 QQPPLPLOCCHLHOEPLVCVPTLKGAKA VKQOIQQQGGGGOGLHETRIYVO 103
QY 51 RAGEIVSCGV 61
DB 104 TATHLPKVCNI 114

RESULT 8
2SS3_ARATH STANDARD; PRT; 164 AA.
AC P15459;
DT 01-APR-1990 (rel. 14, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE 2S seed storage protein 3 precursor (2S albumin storage protein)
DE (NMWU2-2S albumin 3).
GN AT2S3 OR AT4G37160 OR T2A18.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RA Kriebers E., Herdies L., de Clercq A., Seurlinck J., Leemans J.,
RA van Damme J., Segura M., Gheysen G., van Montagu M.,
RA Vandekerckhove J.;
RT "Determination of the processing sites of an Arabidopsis 2S albumin
RT and characterization of the complete gene family.";
RL Plant Physiol. 87:859-866(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RA Conceicao A.D.S., Kriebers E.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schnell C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansoore W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Macho R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Retelert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
```

```
RA Vos P., Hohenseil J., Zimmermann W., Medler H., Ridley P.,
RA Langham S.A., McCullagh B., Billham L., Robben J.,
RA Van der Schueren J., Grymoprez B., Chuang Y.-D., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Delfor E.,
RA Weltzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirse W.,
RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koellter P.,
RA Bernerker S., Hempel S., Feldpausch M., Lamberth J., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., de Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Reckmann T.,
RA Borkova D., Bioecker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argiridou A., Vitale D., Lignori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lechhary A., Aubourg S.,
RA Chefder F., Cooke R., Berger C., Montfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Furnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bleike C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Bedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheel P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kallick J., Graves T., Harmon G., Edwards J.,
RA Latteille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hiller L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Mateo A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999).
RN [4]
RP SEQUENCE OF 103-164 FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Raynal M., Grellet F., Laude M., Meyer Y., Cooke R., Delseny M.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -I- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -I- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M22035; AAA32745.1; -
CC EMBL; 224744; CAA80868.1; -
CC EMBL; AL035680; CAB38846.1; -
CC EMBL; AL161566; CAB79571.1; -
CC EMBL; Z17580; CAA79001.1; -
CC PIR; J01653; NMW03.
CC PIR; S34674; S34674.
CC InterPro; IPR003612; AAI.
CC InterPro; IPR001768; Cereal_1tryp_1tryp_inh.
CC InterPro; IPR000617; Napin.
CC Pfam; PF00234; 1tryp_alpha_1tryp_1.
CC PRINTS; PR00496; NAPIN.
CC PRODOM; PD002498; Napin. 1.
CC SMART; SM00499; AAI; 1.
CC Seed storage protein; Albumin; Signal; Multigene family.
```

```

FT SIGNAL 1 21
FT PROPEP 22 37
FT CHAIN 38 72 2S SEED STORAGE PROTEIN 3 SMALL SUBUNIT
FT PROPEP 73 81 (BY SIMILARITY)
FT CHAIN 82 164 2S SEED STORAGE PROTEIN 3 LARGE SUBUNIT
SQ SEQUENCE 164 AA; 18762 MW; C9EBE6710549F2248 CRC64; (BY SIMILARITY).

Query Match 28.3%; Score 101; DB 1; Length 164;
Best Local Similarity 36.6%; Pred. NO. 0.0008;
Matches 26; Conservative 11; Mismatches 28; Indels 6; Gaps 3

OY 2 QOESEQLQCCNQKQVREDCCECAIKYIAEDQIQGQLHGE-ESERVAQRAGEIVSSCG 60
DB 86 QGGYLLDQCCEMLRQEFPVCVCPFLKQAARAVSLQGQ-HGFPOSRIKYOSAKYLPIKIC 144
OY 61 VR---CMROT 67
DB 145 IQQVGECPEPT 155

RESULT 9
CG2L_LUPAN STANDARD; PRT; 80 AA.
ID CG2L_LUPAN
AC P09931;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Conglutin delta-2 large chain.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxId=3871;
RN [1]
RP SEQUENCE.
RC STRAIN=cv. WHITE;
RA Lillay G.G., Inglis A.S.;
RT "Amino acid sequence of conglutin delta, a sulfur-rich seed protein
of lupinus angustifolius L. Sequence homology with the C-II alpha-
amylase inhibitor from wheat."
RL FEBS Lett. 195:235-241(1986).
RR -I- SUBUNIT: DIMER OF A SMALL CHAIN AND A LARGE CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC CC -I- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR PIR: B23617; B23617.
DR InterPro: IPR003612; AAT.
KW SMART: SM00499; AAT; 1.
RW Seed.
FT FT DISULFID 17 17 INTERCHAIN (WITH C-20 OF SMALL CHAIN)
FT FT DISULFID 18 65 (OR 18).
FT FT DISULFID 29 29 OR 17-65.
FT FT DISULFID 31 73 INTERCHAIN (WITH C-8 OF SMALL CHAIN).
SQ SEQUENCE 80 AA; 9400 MW; 2FA2A76D15A77B6B CRC64;

Query Match 27.5%; Score 98; DB 1; Length 80;
Best Local Similarity 32.9%; Pred. NO. 0.00084;
Matches 23; Conservative 16; Mismatches 23; Indels 8; Gaps 4;

OY 1 QOESEQLQCCNQKQVREDCCECAIKYIAEDQIQGQLHGESERVAGRAGEIVSS-- 57
DB 7 ESESEELDDQCEQNELNSORCGRALQQTIESGS--QCGRROEO-QLEGELEKLKP 62
OY 58 -SCGVRCMRQ 66
DB 63 RICGCGPLRR 72

RESULT 10
```

ID	2SS2_BRANA	STANDARD:	PRT:	178 AA.
AC	P01090.			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	01-AUG-1992 (Rel. 23, Last annotation update)			
DE	Napin 2 precursor (1.7S seed storage protein).			
OS	Brassica napus (Rape).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Brassica.			
OX	NCBI_TaxID=3708;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	EMBL=87308224; PubMed=3624251;			
RA	Josefsson L.-G., Lenman M., Ericson M.L., Rask L.;			
RT	"Structure of a gene encoding the 1.7 S storage protein, napin, from			
RL	Brassica napus."			
RN	J. Biol. Chem. 262:12196-12201 (1987).			
RN	[2]			
RP	REVISIONS.			
RA	Josefsson L.-G.;			
RL	submitted (JUL-1987) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	EMBL=87033665; PubMed=3771543;			
RA	Ericson M.L., Roedin J., Lenman M., Glimelius K., Josefsson L.-G.,			
RA	Rask L.;			
RT	"Structure of the rapeseed 1.7 S storage protein, napin, and its			
RT	precursor."			
RL	J. Biol. Chem. 261:14576-14581 (1986).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. TOWER;			
RX	MEDLINE=84113267; PubMed=6689334;			
RA	Crouch M.L., Tenberge K.M., Simon A.E., Ferl R.;			
RT	"cDNA clones for Brassica napus seed storage proteins: evidence from			
RT	nucleotide sequence analysis that both subunits of napin are cleaved			
RT	from a precursor polypeptide."			
RL	J. Mol. Appl. Genet. 2:273-283 (1983).			
CC	-1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE			
CC	TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING			
CC	ITS MATURATION.			
CC	-1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN			
CC	LINKED BY DISULFIDE BONDS.			
CC	-1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.			
CC	-1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way com-			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; K01545; AAA33006.1; -			
DR	EMBL; J02586; AAA32997.1; -			
DR	EMBL; J02798; AAA87348.1; -			
DR	PIR; A01329; NMRP2			
DR	PIR; A29801; A29801.			
DR	PIR; A25997; A25997.			
DR	InterPro; IPR003612; AAI.			
DR	InterPro; IPR001768; Cereal_tryp-amy1_inh.			
DR	InterPro; IPR000617; Napin.			
DR	Pfam; PF00234; tryp_alpha_amy1.1.			
DR	PRINTS; PR00496; NAPIN.			
DR	PRODom; PD002496; Napin.1.			
DR	SMART; SM00499; AAI.1.			
DR	Seed storage proteoln; Signal; Multigene family.			
FT	SIGNAL	1	21	
FT	PROPEP	22	38	
FT	CHAIN	39	74	SMALL CHAIN.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:45:36 : Search time 77.42 Seconds
(without alignments)
151.946 Million cell updates/sec

Title: US-09-913-351-3
Perfect score: 357
Sequence: 1 QQESQQLQCCNQVKNQVRD.....AQRAGETVSSGCVACMRQTR 68

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp rodent:*
12: sp.virus:*
13: sp vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	39.5	139	10	P93198
2	132.5	37.1	323	10	Q39928
3	128	35.9	153	10	Q9AUD1
4	124	34.7	65	10	Q9S872
5	122	34.2	66	10	Q9S870
6	122	34.2	141	10	Q39649
7	112	31.4	64	10	Q9AVK2
8	109	30.5	161	10	Q81410
9	107.5	30.1	165	10	Q40850
10	102	28.6	153	10	Q99235
11	99	27.7	172	10	Q40997
12	97	27.2	167	10	Q81411
13	96	26.3	184	10	Q42444
14	95.5	26.8	91	10	Q9S9E5
15	95.5	26.8	162	10	Q64929
16	95	26.6	162	10	Q40998

17	92	25.8	165	10	Q9FH31
18	91	25.5	86	10	Q9S9F0
19	91	25.5	88	10	Q9S9E7
20	91	25.5	125	10	P80208
21	91	25.5	178	10	Q42413
22	91	25.5	178	10	Q42473
23	91	25.5	178	10	Q96339
24	91	25.5	180	10	Q42469
25	90.5	25.4	91	10	Q9S9E5
26	90.5	25.4	155	10	Q9ZN24
27	90	25.2	173	10	Q64930
28	89.5	25.1	148	10	Q9XHP1
29	89	24.9	162	10	Q64932
30	89	24.9	190	10	Q40995
31	88.5	24.8	173	10	Q9Z8R1
32	88	24.6	139	10	Q39795
33	88	24.6	285	10	Q41603
34	87.5	24.5	152	10	Q41168
35	87.5	24.5	156	10	Q941R0
36	87	24.4	88	10	Q9S9E9
37	87	24.4	88	10	Q9S9E8
38	86	24.1	178	10	Q39344
39	85.5	23.9	146	10	Q9JRC2
40	85.5	23.9	170	10	Q64931
41	85.5	23.9	174	10	Q41169
42	85.5	23.9	318	10	Q41545
43	85	23.8	155	10	Q41167
44	84.5	23.7	173	10	Q81412
45	84	23.5	178	10	Q42490

ALIGNMENTS

RESULT 1
P93198 PRELIMINARY: PRT: 139 AA.
AC P93198;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALBUMIN SEED STORAGE PROTEIN (FRAGMENT).
OS Juglans regia (English walnut).
OC Eukaryota; Viridiplantae; streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fagales; Juglandaceae; Juglans.
OX NCBI_Taxid=51240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNLAND;
RA Teuber S.S., Dandekar A.M., Peterson W.R.;
RT "Juglans regia 2S albumin seed storage protein precursor."
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U66866; AAB41308.1; -.
DR InterPro: IPR001768; Cereal_1.
DR InterPro: IPR001768; Cereal_1.
DR InterPro: IPR000480; Glutelin.
DR Pfam: PF00234; tryp_alpha_1.
DR PRINTS: PR00211; GLUTELIN.
DR SMART: SM00499; AAI; 1.
FT NON_TER
SQ SEQUENCE 139 AA: 16373 MW: 02D0E55E67164F23 CRC64;

Query Match 39.5%; Score 141; DB 10; Length 139;
Best Local Similarity 35.4%; Pred. No. 1.9e-08;
Matches 23; Conservative 15; Mismatches 27; Indels 0; Gaps 0;

QY 1 QQESQQLQCCNQVKNQVRDCECEAIKYIAEQIQGGULHGESPRVQARAGETVSSG 60
DB 67 EENQORHFQCCQOLSSQMBQCCCEGLRVVRRQDQDQGGILRGEEMEVQASADLPNECG 126
QY 61 VRCMR 65


```

Query Match
Best Local Similarity 30.5%; Score 109; DB 10; Length 161;
Matches 22; Conservative 21; Mismatches 20; Indels 18; Gaps 2;

QY 1 QOSESQLOQCCNOVKYRDECCCEAIXYIAEDQI-----QOQOLHGESESR 47
DB 66 ERRRRQPSERCCCELIQRMSPOCRCAIOTLEDYFMDSODGAPLNRGCGRMEEER 125
QY 48 VAORAGEIVSSGV-----RC 63
DB 126 VLRRAEELPNTCNVQSPRR 146

RESULT 9
Q40850 PRELIMINARY; PRT; 165 AA.
AC Q40850;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SEED STORAGE PROTEIN.
GN EMB25.
OS Picea glauca (White spruce).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
OX NCBI_TaxID=3330;
RN [1]
RP SEQUENCE FROM N.A.
RT Dong J.-Z., Dunstan D.I.;
RL "Gene expression during somatic embryogenesis."
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: I47745; AAB01560.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_amyL_inh.
DR Pfam: PF00234; tryp_alpha_amyL; 1.
DR SMART: SM00499; AAI; 1.
SQ SEQUENCE 165 AA; 19396 MW; 69CD273982CA0D20 CRC64;

Query Match
Best Local Similarity 30.1%; Score 107.5; DB 10; Length 165;
Matches 21; Conservative 20; Mismatches 21; Indels 17; Gaps 1;

QY 1 QOSESQLOQCCNOVKYRDECCCEAIXYIAED-----QOQOLHGE 43
DB 66 ERRRRQPSERCCCELIQRMSPOCRCAIOTLEDYFMDSODGAPLNRGCGRM 125
QY 44 ESERVARAGEIVSSGV 62
DB 126 EEEVYRRAEELPNTCNVR 144

RESULT 10
Q409235 PRELIMINARY; PRT; 153 AA.
AC Q409235;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CONGLUTIN DELTA.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicotyledons; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3871;
RN [1]
RP SEQUENCE FROM N.A.
RT STRAIN=UNICROP; TISSUE=COTYLEDON;
RX MEDLINE=91355912; PubMed=2103479;
RA Gayler K.R., Kollivas S., Macfarlane A.J., Lilley G.G., Baldi M.,
Blagrove R.J., Johnson E.D.;

```

```

RT "Biosynthesis, CDNA and amino acid sequences of a precursor of
RT conglutin delta, a sulphur-rich protein from Lupinus angustifolius.";
RL Plant Mol. Biol. 15:879-893(1990).
DR EMBL: X53523; CA37598.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_amyL_inh.
DR Pfam: PF00234; tryp_alpha_amyL; 1.
DR SMART: SM00499; AAI; 1.
SQ SEQUENCE 153 AA; 17785 MW; A45B702535F0769 CRC64;

Query Match
Best Local Similarity 28.6%; Score 102; DB 10; Length 153;
Matches 19; Conservative 21; Mismatches 23; Indels 2; Gaps 2;

QY 3 QESQLOQCCNOVKYRDECCCEAIXYIAEDQI-----QOQOLHGESESRVORAGEIVSSGV 61
DB 82 ESEELDQCCCELIQRMSPOCRCAIOTLEDYFMDSODGAPLNRGCGRM 140
QY 62 RCMRQ 66
DB 141 GPLRR 145

RESULT 11
Q40997 PRELIMINARY; PRT; 172 AA.
AC Q40997;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ALBUMIN 3 (FRAGMENT).
OS Pinus strobus (Eastern white pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3348;
RN [1]
RP SEQUENCE FROM N.A.
RT Rice J., Kamalay J.C.;
RL "A family of pine seed proteins related to 2S albumins.";
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL: X62435; CAA44300.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_amyL_inh.
DR Pfam: PF00234; tryp_alpha_amyL; 1.
DR SMART: SM00499; AAI; 1.
FT NON_TER 1
SQ SEQUENCE 172 AA; 19858 MW; 2AE6FC9B210C7BF5 CRC64;

Query Match
Best Local Similarity 27.7%; Score 99; DB 10; Length 172;
Matches 19; Conservative 17; Mismatches 26; Indels 22; Gaps 1;

QY 1 QOSESQLOQCCNOVKYRDECCCEAIXYIAEDQI-----QO 38
DB 71 QRRRQPSERCCCELIQRMSPOCRCAIOTLEDYFMDSODGAPLNRGCGRM 130
QY 39 QLHGESESRVORAGEIVSSGV 62
DB 131 RGRREEEVEAERAGELPDRCNVR 154

RESULT 12
Q81411 PRELIMINARY; PRT; 167 AA.
AC Q81411;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 2S ALBUMIN-LIKE PROTEIN.
OS Picea glauca (White spruce).

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
OX NCBI_TaxID=3330;
RN [1]
RP SEQUENCE FROM N.A.
RA Dong J.-Z., Dunstan D.I.;
RT "Picea glauca 2S albumin-like cDNA (EMB37).";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF074938; AAC34614.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1tryp_aml_1nh.
DR Pfam; PF00234; tryp_alpha_aml1; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 167 AA; 19726 MW; 673422192EF2A212 CRC64;

Query Match 27.8%; Score 97; DB 10; Length 167;
Best Local Similarity 24.1%; Pred. No. 0.002;
Matches 21; Conservative 19; Mismatches 23; Indels 24; Gaps 2;

QY 1 QOQESQOQQCCNOVKQVRDECCCEAIKTAEDQIQ-----QGQLH 41
DB 66 EERRDPSERCCELRMSQCRCAQRTLEDVFNMSQSDSAPLNRRRRRGRGQR 125
QY 42 GESESVARAGEIVSSCGV-----RC 63
DB 126 EMEEEVVRAREELPNTCNVLQSPRC 152

RESULT 13

042444 PRELIMINARY; PRT; 184 AA.
AC 042444;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NAPIR.
OS Brassica campestris (Field mustard).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3711;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. R500;
RA Kridl J.C., McCarter D.W., Rose R.E., Scherer D.E., Knutzon D.S.,
RA Radke S.E., Knauf V.C.;
RL Seed sci. Res. 1:209-219(1991).
DR EMBL; M64632; AAA32999.1; -.
DR EMBL; M64631; AAA32998.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1tryp_aml_1nh.
DR InterPro; IPR000617; Napiin.
DR Pfam; PF00234; tryp_alpha_aml1; 1.
DR PRINTS; PR00496; NAPIN.
DR Prodom; PD002498; Napiin; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 184 AA; 20865 MW; BC2E46C039FCE9A2 CRC64;

Query Match 26.9%; Score 96; DB 10; Length 184;
Best Local Similarity 35.5%; Pred. No. 0.0029;
Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;
QY 1 QOQESQO---LQCCNOVKQVRDECCCEAIK---YIAEDQI--QOG-QLHGESE----- 46
DB 92 QOQGPQGRPLLQCCNELHQEELVCVPTLKASKAVKQVIRQOQGGQOQGGQMOQVYS 151
QY 47 RVAQRAGEIVSSCGVR 62
DB 152 RIVQTATHLPRACNIR 167

RESULT 14

09S9E6 PRELIMINARY; PRT; 91 AA.
AC 09S9E6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE NAPIR LARGE CHAIN L2B-CALMODULIN ANTAGONIST/CALCIUM-DEPENDENT PROTEIN
DE KINASE SUBSTRATE.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE.

RX MEDLINE=96283791; PubMed=8679671;
RA Neumann G.M., Condron R., Thomas I., Polya G.M.;
RT "Purification and sequencing of multiple forms of Brassica napus seed
RT napin large chains that are calmodulin antagonists and substrates for
RT plant calcium-dependent protein kinase.";
RL Biochim. Biophys. Acta 1295:34-43(1996).
DR InterPro; IPR003612; AAI.
DR InterPro; IPR000617; Napiin.
DR PRINTS; PR00496; NAPIN.
DR Prodom; PD002498; Napiin; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 91 AA; 10276 MW; 300CDA9453F68667 CRC64;

Query Match 26.8%; Score 95.5; DB 10; Length 91;
Best Local Similarity 31.9%; Pred. No. 0.0015;
Matches 23; Conservative 10; Mismatches 28; Indels 11; Gaps 2;

QY 2 QOQESQOQQCCNOVKQVRDECCCEAIKTAEDQIQGGQLHGESE-----RVAQ 50
DB 5 QGRPPLQCCNELHQEELVCVPTLKASKAVKQVIRQOQGGQOQGGQLQVIRIQ 64
QY 51 RAGEIVSSCGVR 62
DB 65 TATHLPRVCNIR 76

RESULT 15

064929 PRELIMINARY; PRT; 162 AA.
AC 064929;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 2S SEED STORAGE PROTEIN.
GN PM2S1.
OS Pseudotsuga menziesii (Douglas-fir).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pseudotsuga.
OX NCBI_TaxID=3357;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRANCO;
RA Chatthaal M., Misra S.;
RX MEDLINE=98381325; PubMed=9715536;
RT "Sequence and expression of embryogenesis-specific cDNAs encoding 2S
RT seed storage proteins in Pseudotsuga menziesii [Mrb.] Franco.";
RL Planta 206:138-145(1998).
DR EMBL; AF029970; AAC26998.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1tryp_aml_1nh.
DR Pfam; PF00234; tryp_alpha_aml1; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 162 AA; 18705 MW; 6139481B1D31FDC2 CRC64;

Query Match 26.8%; Score 95.5; DB 10; Length 162;

Best Local Similarity 25.6%; Pred. No. 0.0028;
Matches 21; Conservative 21; Mismatches 21; Indels 19; Gaps 4;

QY 1 000ESQQLQCCCNQVQVDEOCCEAIKXI-----AEDQIQ--GQLHG---EESF 46
Db 66 ERRREDPPSSCCNELERMSPPQCRCPAIOQVLDOSASFMDSEDLNQRGRREGRRREQ 125

QY 47 RVAQRAGEIVSSCGV-----RC 63
Db 126 EMARAAVLPDTGNVQESPPRC 147

Search completed: September 9, 2002, 12:45:37
Job time: 294 sec

PT Ricinus communis -
XX
PS
XX Claim 3; Page 30; 38pp; Spanish.
XX
CC The invention relates to a novel glycoconjugate (A) comprising a
CC 50-250 kDa polysaccharide (PS), containing one phosphate group per
CC 5-25 monosaccharide units, and a polypeptide (PP) having a particular
CC consensus sequence. In PS, at least 40% of the monosaccharide residues
CC are mannose and the remainder are glucose and/or galactose. The main
CC chain comprises 1-6 links, with less than 60% 1,2-branching. The PP has
CC the consensus sequence:
CC Zacczbc(QEKK)z(z')(LTVW)zcC(z'')(OEH)(LV)zCcCzC22aGzCz(VILM)zCfcZg
CC where Z = any amino acid (aa); Z' = hydrophobic aa; Z'' = hydrophilic aa;
CC a = 3-48; b = 9-13; c = 15-39; d = 13-56; e = 15-26; f = 1-8; g = 1-11;
CC the parentheses indicate the preferred sequence. This sequence
CC corresponds to amino acids 157-190 of the Ricinus communis (castor bean)
CC 2S albumin storage protein which is used to generate the glycoconjugate
CC of the invention, with a polysaccharide from candida utilis. (A) modifies
CC the immune response in humans and animals, especially it reduces the
CC amount of tumour necrosis factor (TNF) produced and stimulates the
CC mononuclear-phagocyte system and expands the granulocyte-macrophage
CC compartment. (A) are used for treating disorders of the immune system,
CC e.g. infections and tumours. (A) do not interfere with metabolic
CC processes in the liver and are active when given orally. PP significantly
CC increases the antigenic response to weakly immunogenic PS, including
CC induction of a T-dependent response.
XX
XX Sequence 34 AA:
XX

PI Knauf VC, Kridl JC;
XX
XX WPI: 2001-564354/63.
XX
XX Obtaining a plant that produces a seed with a modified phenotype or
XX altering a seed phenotype, comprises transforming a plant cell with a
XX DNA construct consisting of operably linked components in the direction
XX of transcription -
XX
XX Example 9: Fig 6, 68pp; English.
XX
XX The invention provides a method for obtaining a plant which produces
XX at least one seed having a modified phenotype. The method involves
XX transforming a host plant cell with a DNA construct which consists of
XX operably linked components in the direction of transcription, a promoter
XX region from a Brassica plant gene, a DNA sequence of interest other than
XX the native coding sequence, and a transcription termination region. The
XX method is useful for obtaining plants having modified phenotype or for
XX altering the phenotype of a plant seed or tissue. The DNA constructs are
XX used in manipulating plant cells to provide for regulated transcription,
XX such as light inducible transcription, in a plant tissue or plant part of
XX interest at particular stages of plant growth or in response to external
XX control. These constructs are also used for modulation of expression of
XX endogenous products as well as production of exogenous products in the
XX seed. Sequences AAB5941-47 represent fragments of storage proteins used
XX in comparison studies with the storage protein 2A11.
XX
XX Sequence 17 AA:
XX

Query Match	100.0%	Score 183	DB 21	Length 34
Best Local Similarity	100.0%	Pred. No. 6	Be-18	
Matches 34	Conservative 0	Mismatches 0	Indels 0	Gaps 0
OY	1	PSQGCRCQIQDEQNLRQCQETIKKQVSGQGR	34	
db	1	psggcrrgqiqeqnllrqqeyikqvsqgqrr	34	

Query Match	42.9%	Score 78.5	DB 22	Length 17
Best Local Similarity	94.4%	Pred. No. 0.00522		
Matches 17; Conservative	0;	Mismatches 0;	Indels 1;	Gaps 1;
OY	13	QONLRKQCEYIKQVSGQ	30	
db	1	qgnlrqcgeyl-qgvsqg	17	

RESULT	2
AAB85946	
ID	AAB85946 standard; peptide; 17 AA.
XX	
AC	AAB85946;
XX	
DT	30-NOV-2001 (first entry)
XX	
DE	Castor bean 2S small subunit protein fragment

	RESULT	3
XX	AA62762	
ID	AA62762	standard; Protein; 110 AA.
XX	AC	
XX	AA62762;	
XX	07-JUL-1995	(first entry)
XX	Synthetic 12 Kd precursor protein of Brazil nut.	

KW	Napin gene; promoter; Brassica; phenotype; transcription; EA9; 3h11,
KM	2A11; alpha-amylase.
XX	
OS	Richius communis.
XX	
PN	US6281410-B1.
XX	
PD	28-AUG-2001.
XX	
PE	15-JAN-1999; 99US-0232861.
XX	
PR	29-APR-1988; 88US-0188361
PR	02-NOV-1988; 88US-0267683.
PR	10-AUG-1993; 93US-0105852.
PR	07-JUN-1995; 95US-0484941.
PR	07-MAR-1997; 97US-0812665.
PR	31-JUL-1986; 86US-0891529.
PR	26-MAY-1987; 87US-0054369.
PR	28-JUL-1987; 87US-0078538.
PR	25-JAN-1988; 88US-0147781.
PR	15-MAR-1988; 88US-0168190.

KW	Stabilisation; toxic; peroxisome; yeast; food additive; brazil nu
XX	
OS	Synthetic.
XX	
PN	W09424289-A.
XX	
PD	27-OCT-1994.
XX	
PF	19-APR-1994; 94WO-FR00438.
XX	
PR	19-APR-1993; 93FR-0004583.
XX	
PA	(EURO-) EUROLYSINE.
XX	
XX	
PI	Ito H, Labat N, Nicaud J, Pardo D, Raynal A, Sugimoto S;
XX	
DR	WPI: 1994-341868/42.
XX	
PT	N-PSDB; AAQ73069.
XX	
PT	Improving stability of protein or peptide in yeast - by
PT	expressing with a peroxisomal targeting sequence, allowing
PT	

PA (CALJ) CALGENE LLC.
XX

PT	useful as food additives	
XX		

AAW86271
ID AAW86271 standard; Protein; 111 AA.
XX
AC AAW86271;
XX
DT 01-MAR-1999 (first entry)
XX
DE Modified Brazil nut 2S albumin gene BN153KW amino acid sequence.
XX
KW Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KW BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN153W;
KW chimeric; dietary protein; essential amino acid.
XX
OS Bertholletia excelsa.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 4 /label= E4K
FT /note= "wild-type Glu is replaced with Lys"
FT
FT Misc-difference 6 /label= R6K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 11 /label= R11K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 16 /label= S16K
FT /note= "wild-type Ser is replaced with Lys"
FT
FT Misc-difference 19 /label= R19K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 23 /label= R23K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 36 /label= R36K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 37 /label= R37K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 41 /label= P41K
FT /note= "wild-type Pro is replaced with Lys"
FT
FT Misc-difference 42 /label= W42H
FT /note= "wild-type Trp is replaced with His"
FT
FT Misc-difference 58 /label= R58K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 63 /label= R63K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 68 /label= R68K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 77 /label= R77K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 82 /label= R82K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 83 /label= R83K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 86 /label= R86K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 94 /label= R94K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 101 /label= R101K
FT

FT /note= "wild-type Arg is replaced with Lys"
XX
XX WO9845458-A1.
XX
XX 15-OCT-1998.
XX
XX
XX
XX 06-APR-1998; 98WO-US06673.
XX
XX 08-APR-1997; 97US-0042827.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX
XX Gutteridge S;
XX
XX WPI; 1998-609902/51.
XX
XX N-PSDB; AAW80251.
XX
XX New modified Brazil Nut 2S albumin proteins - useful to provide
PT transgenic seed crops whose proteins have increased levels of
PT essential amino acids
XX
XX Claim 6; Fig 8; 80pp; English.
XX
XX
XX The invention relates to modified Brazil Nut 2S albumin seed storage
CC proteins and nucleic acids encoding the modified proteins. The modified
CC Brazil nut 2S albumin proteins are rich in essential amino acids and the
CC encoding genes selected from a group comprising of BNCNSS, BN11, BN15,
CC BN17, BN18, BN19, BN153KW, AT2S1BN15, AT2S1BN153W genes.
CC Chimeric genes comprising a nucleic acid fragment encoding a modified
CC Brazil nut 2S protein operably linked to suitable regulatory sequences,
CC integrated into the genome of a transformed plant can be used in
CC methods for increasing the essential amino acid content of seeds. The
CC products can be used to provide a human and animal dietary protein in
CC seed crop plants that is relatively high in essential amino acid content.
CC The present sequence represents the amino acid sequence of a modified
CC Brazil Nut 2S albumin gene BN153KW.
XX
SO Sequence 111 AA:

Query Match 37.2%; Score 68; DB 19; Length 111;
Best Local Similarity 44.0%; Pred. No. 0.11;
Matches 11; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 3 QGCGRGIOEQNLRQCQERYIKQOV 27
DB 2 gekckegmqkqkmlkhckymkqgm 26

RESULT 9
AAW86265
ID AAW86265 standard; Protein; 111 AA.
XX
AC AAW86265;
XX
DT 01-MAR-1999 (first entry)
XX
XX Modified Brazil nut 2S albumin gene BNCNSS amino acid sequence.
DE
XX
KW Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KW BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN153W;
KW chimeric; dietary protein; essential amino acid.
XX
OS Bertholletia excelsa.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 37 /label= R37K
FT /note= "wild-type Arg is replaced with Lys"
FT
FT Misc-difference 58 /label= R58K
FT /note= "wild-type Arg is replaced with Lys"
FT

DR	WPI; 1994-166646/20.
DR	N-PsDB; AAQ65428.
PT	DNA encoding plant acyl carrier protein - used to enhance prodn.
PT	of seed oil and to modulate the fatty acid compsn. of the oil
XX	
PS	Disclosure; Fig 4E-4G; 28pp; English.
CC	AAFS4981 shows a Brassica napin protein, generated from lambda CGN1-2
CC	clone. The napin gene promoter is regulated during seed maturation.
CC	This can be used to control the expression of an ACP in seed to
CC	enhance the production of seed oil and modulate the fatty acid
CC	composition. The DNA can also be used to prepare probes.
XX	
SQ	Sequence 184 AA;
Query Match	37.2%; Score 68; DB 15; Length 184;
Best Local Similarity	41.4%; Pred. NO. 0.18;
Matches	12; Conservative 8; Mismatches 7; Indels 2; Gaps 1;
QY	6 CNGQIOEQONLRQCCEXIKQGV--SGGP 32 : : : : : : : Db 47 crkefqgaqhlkacqgwlnkqamsgsgsp 75
RESULT 12	
ID	AAR76694
AC	AAR76694 standard; Protein; 184 AA.
XX	
AC	AAR76694;
XX	
DT	04-NOV-1995 (first entry)
DE	B. campestris napin.
XX	
KM	Napin; promoter; seed-specific gene transcription; oilseed;
KW	transgenic plant.
XX	
OS	Brassica campestris.
XX	
PN	US5420034-A.
XX	
PD	30-MAY-1995.
PF	
PE	31-JUL-1986; 86US-0891529.
XX	
PR	09-JUL-1990; 90US-0550804.
PR	31-JUL-1986; 86US-0891529.
PR	28-JUL-1987; 87US-0078538.
PR	25-JAN-1988; 88US-0147781.
PR	08-AUG-1991; 91US-0742834.
XA	(CALJ) CALGENE INC.
XX	
PI	Knauf VC, Kridl JC;
XX	
DR	WPI; 1995-206250/27.
DR	N-PsDB; AAQ94586.
XX	
PT	DNA providing seed specific gene transcription - includes a
PT	transcription initiator from a gene preferentially expressed in
PT	seed cells, e.g. the napin gene, also related expression cassettes
PT	and Brassica plants.
XX	
PS	Disclosure; Fig. 2A-J; 40pp; English.
XX	
CC	Clone lambda CGN1-2 was isolated by screening a partial genomic
CC	library of B. campestris DNA for napin-encoding clones. The entire
CC	CDS as well as extensive 5' and 3' sequences of the napin gene were
CC	sequenced (AAQ94586). An expression cassette was constructed from
CC	the 3'- and 5'-ends of the lambda CGN1-2 napin gene.
XX	

```

SO      Sequence      184 AA;

Query Match      37.2%, Score 68; DB 16; Length 184;
Best Local Similarity 41.4%; Pred.No. 0.18;
Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 1;

OY      6 CRGQIQEQGNLRQCCQERYIKQV--SGQSP 32
      || : | : | : | : | : | : | : | : |
DB      47 CRKEFGQGHKACQGWLHKQAMGSGSP 75

RESULT 13
AAW15415
ID      AAW15415 standard; Protein: 184 AA.
AC      AAW15415;
XX
XX      05-JUN-1997 (first entry)
XX
XX      Napin encoded by clone gamma-ACGN1-2.
XX
XX      Plant seed cell; napin gene; acyl carrier protein gene; EA9 gene;
KW      Brassica; fatty acid; rat thioesterase; acyl ACP;
KW      phospholipid acyl desaturase.
XX
XX      Brassica campestris.
XX
XX      US5608152-A.
XX
XX      04-MAR-1997.
XX
XX      31-JUL-1986; 86US-0891529.
XX
XX      09-JUL-1990; 90US-0550804.
XX      31-JUL-1986; 86US-0891529.
XX      28-JUL-1987; 87US-0078538.
XX      25-JAN-1988; 88US-0147781.
XX      08-AUG-1991; 91US-0742834.
XX      30-MAY-1995; 95US-0453924.
XX
XX      (CALJ ) CALGENE INC.
XX
XX      Knaut VC, Kridl JC;
XX
XX      WPI: 1997-164560/15.
XX      N-PSDB: AAT65475.
XX
XX      Transgenic Brassica plant and seeds - contain DNA construct for seed
XX      PT specific expression of heterologous proteins
XX
XX      Example 2; Column 23-28; 48pp: English.
XX
XX      The sequences given in AAW15415-17 represent proteins encoded by DNA
XX      fragments which were used in the construction of the DNA construct
XX      CC of the invention. The DNA construct comprises, in the 5' to 3'
XX      CC direction of transcription;
XX      CC (a) a transcriptional initiation region from a gene that encodes a
XX      CC product preferentially expressed in a plant seed cell as compared to
XX      CC other plant cells;
XX      CC (b) a DNA sequence of interest other than the native coding sequence
XX      CC of the gene; and
XX      CC (c) a transcriptional termination region;
XX      CC where the gene is a napin gene, an acyl carrier protein gene or an EA9
XX      CC gene. This construct may be used in the production of a Brassica plant
XX      CC which may be modified such that heterologous proteins are produced in
XX      CC the seeds, such as fatty acids, rat thioesterase, acyl ACP or
XX      CC phospholipid acyl desaturase.
XX
XX      Sequence      184 AA;

Query Match      37.2%; Score 68; DB 18; Length 184;

```


Mon Sep 9 13:14:50 2002

us-09-913-351-4.rag

Page 9

XX. Sequence 184 AA;

Query Match 37.2%; Score 68; DB 19; Length 184;
Best Local Similarity 41.4%; Pred. NO. 0.18;
Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 1.

Qy 6 CRGQIQEQNLRCQCEYIKQV--SGQGP 32
Db 47 crkefqgaqlkacqqlhkgamqsgsp 75

Search completed: September 9, 2002, 12:42:39
Job time: 191 sec

THIS PAGE BLANK (USPTO)

Db 38 NOPLCRPOFOHQHILKACQRYIRRAQRG 67

RESULT 2

US-08-453-924-3

Sequence 3, Application US/08453924

Patent No. 5608152

GENERAL INFORMATION:

APPLICANT: Kridl, Jean C.

APPLICANT: Knaut, Vlc C.

TITLE OF INVENTION: Seed-specific Transcriptional Regulation

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weill, Gotshal & Manges

STREET: 2882 Sand Hill Road, Ste. 280

CITY: Menlo Park

STATE: CA

COUNTRY: U.S.A.

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/453,924

FILING DATE: 30-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/742,834

FILING DATE: 08-AUGUST-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/550,804

FILING DATE: 09-JULY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/147,781

FILING DATE: 25-JANUARY-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/078,538

FILING DATE: 28-JULY-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 06/891,529

FILING DATE: 31-JULY-1986

ATTORNEY/AGENT INFORMATION:

NAME: Rae-Ventler, Barbara

REGISTRATION NUMBER: 32,750

REFERENCE/DOCKET NUMBER: CGNE-037/04US

TELEPHONE: (415) 926-6200

TELEFAX: (415) 854-3713

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 184 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-453-924-3

Query Match 37.2%; Score 68; DB 1; Length 184;

Best Local Similarity 41.4%; Pred. No. 0.071;

Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 1;

Db 47 CRKEFOQAOHLKACQWILKQAMQSGSG 75

US-08-670-186-6

Sequence 6, Application US/08670186

Patent No. 5859343

GENERAL INFORMATION:

APPLICANT: SUN, SAMUEL S.M.

APPLICANT: XIONG, LIWEN

APPLICANT: HU, ZHONG

APPLICANT: CHEN, HANG

TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500

CITY: WASHINGTON

STATE: DC

COUNTRY: USA

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/670,186

FILING DATE: 21-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 23461-20007,00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 822-0168

TELEX: 90-4030 MRSNFOERSWSH

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 158 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-670-186-6

Query Match 35.0%; Score 64; DB 2; Length 158;

Best Local Similarity 36.7%; Pred. No. 0.2;

Matches 11; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Db 38 NOPLCRPOFOHQHILKACQRYIRRAQRG 67

US-08-618-911-4

Sequence 4, Application US/08618911

Patent No. 5850016

GENERAL INFORMATION:

APPLICANT: Jung, Rudolf

APPLICANT: Hastings, Craig

APPLICANT: Coughlan, Sean

APPLICANT: Hu, David

TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN

TITLE OF INVENTION: SEEDS

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.

STREET: 700 Capital Square, 400 Locust Street

CITY: Des Moines

STATE: Iowa

COUNTRY: USA

ZIP: 50309

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:


```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/07/955,905A
:
: FILING DATE: 21-Jan-1993
:
: CLASSIFICATION: 435
:
: INFORMATION FOR SEQ ID NO: 2:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 566 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-07-955-905A-2

```

Query Match	30.6%;	Score 56;	DB 1;	Length 566;
Best Local Similarity	37.1%;	Pred. NO. 8.5;		
Matches 13; Conservative	5;	Mismatches 9;	Indels 8;	Gaps 1;

```

QY      4 QGCRGQIQEQN-----LRQCQYIKQVSG 30
          | | | | | | | | | | | | | |
Db      86 QQCGRGQEQQQGQREQQCCRKQWQYKQERGE 120

```

```

RESULT      8
US-07-955-905A-22
: Sequence 22, Application US/07955905A
: Patent No. 5770433
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
: TITLE OF INVENTION: PRECURSOR
: NUMBER OF SEQUENCES: 28
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/955,905A
: FILING DATE: 21-JAN-1993
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 566 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Theobroma cacao
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..566
: OTHER INFORMATION: /note="67 kD Precursor Protein"
: US-07-955-905A-22

```

QY	4	QCGRCGQIDQDN-----LRCCGCTKQGVSG	30
		: : : : : : : : : : : :	
b	86	QCGRCGQIDQDN-----LRCCGCTKQGVSG	120

Query Match 30.6%; Score 56; DB 1; Length 566;
 Best Local Similarity 37.1%; Pred. No. 8.5;
 Matches 13; Conservative 5; Mismatches 9; Indels 8; Gaps 1.

RESULT 9
US-08-618-911-6
; Sequence 6, Application US/08618911
; Patent No. 5850016
; GENERAL INFORMATION:
; APPLICANT: Jung, Rudolf
; APPLICANT: Hastings, Craig

```

1  APPLICANT: Coughlan, Sean
2  APPLICANT: Hu, David
3  TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS, IN
4  TITLE OF INVENTION: SEEDS
5  NUMBER OF SEQUENCES: 13
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
8  STREET: 700 Capital Square, 400 Locust Street
9  CITY: Des Moines
10 STATE: Iowa
11 COUNTRY: USA
12 ZIP: 50309
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patent In Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/618,911
20 FILING DATE: Concurrently herewith
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Simon, Soma
23 REGISTRATION NUMBER: 37,444
24 REFERENCE/DOCKET NUMBER: 365-US
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: (515) 248-4800
27 TELEFAX: (515) 248-4844
28 INFORMATION FOR SEQ ID NO: 6:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 158 amino acids
31 TYPE: amino acid
32 TOPOLOGY: linear
33 MOLECULE TYPE: protein
34 US-08-618-911-6

```

Query Match	29.8%;	Score 54.5;	DB 2;	Length 158;
Best Local Similarity	34.5%;	Pred. NO. 3.6;		
Matches 10;	Conservative 9;	Mismatches 9;	Indels 1;	Gaps 1;

```

QY      3 QQCGRGQIQEQONLRQCCQEQYIKQVSGG 31
      | | | : : | | : : : : : | |
DB     28 QDSCRQL-KGVNLTPECKHIMEKIQGRG 55

```

RESULT 10
US-08-670-186-2
Sequence 2, Application US/08670186
Patent No. 5859343
GENERAL INFORMATION:
APPLICANT: SUN, SAMUEL S.M.
APPLICANT: XIONG, LIMEN
APPLICANT: HU, ZHONG
APPLICANT: CHEN, HANG
TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,186
FILING DATE: 21-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

TELEPHONE: 617-248-7000
 TELEFAX: 617-248-7100
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 542 amino acids
 TYPE: amino acid
 TOPOLOGY: 1 linear
 MOLECULAR TYPE: peptide
 ORGANISM: Saccharomyces cerevisiae
 INDIVIDUAL SOURCE:
 CLONED FROM: fractional yeast SSNP (fysSRP) (predicted)
 US-08-328-809-8

Query Match 26.2%; Score 48; DB 1; Length 542;
 Basic Local Similarity 24.0%; Pred. No. 92;
 Matched 12; Conservative 8; Mismatches 6; Indels 24; Gaps 1;
 Q: QIQEOONEROCCEYIKO-----QVSGGPPR 34
 I: I I I I I
 D: Q339 QIQEOONEROCCEYIKO-----QVSGGPPR 308

Search completed: September 9, 2002, 12:44:13
 Job time: 235 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:43:30 ; Search time 45.25 Seconds
(without alignments)
72.200 Million cell updates/sec

Title: US-09-913-351-4

Perfect score: 183

Sequence: 1 PSQGGCRGQIQEQNLRQCEYIKQVSGGPPRR 34

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	258	1 RZCS	2S seed storage pr
2	82	44.8	106	2 S26636	napin n1a - rape
3	80	43.7	110	2 S20350	napin n1a - rape
4	76.5	41.8	154	2 S14947	2S albumin - Brazi
5	74	40.4	139	2 T09878	albumin 2S storage
6	73	39.9	32	2 B59346	seed storage prote
7	71	38.8	28	2 A25802	2S seed storage pr
8	71	38.8	37	2 S35591	major allergen sma
9	71	38.8	39	2 S01791	allergen sin a i s
10	71	38.8	127	2 S65447	allergen sin a i s
11	71	38.8	145	2 PC1246	sin a i allergen i
12	71	38.8	145	2 S65478	allergen sin a i
13	71	38.8	145	2 S65481	allergen sin a i
14	71	38.8	145	2 S65482	allergen sin a i
15	71	38.8	145	2 S65480	allergen sin a i
16	71	38.8	145	2 S65479	allergen sin a i
17	71	38.8	152	2 PS0427	allergen sin a i
18	71	38.8	174	2 PS0425	napin AH1 precursor
19	71	38.8	180	2 S52025	napin BA3 precursor
20	70	38.3	139	2 T09850	napin (clones BMN
21	70	38.3	146	2 S14946	albumin 2S storage
22	69	37.7	104	2 S48176	2S seed storage pr
23	69	37.7	124	2 A37931	mablinin I-1 - Yun
24	69	37.7	145	2 PC1247	napin - Swedish tu
25	69	37.7	178	1 NMRP2	sin a i allergen 2
26	69	37.7	178	1 S07828	napin 2 precursor
27	69	37.7	178	1 A25997	napin B - rape
28	69	37.7	178	2 S25137	napin precursor (n
29	68	37.2	178	2 S15382	2S storage protein
					napin napb - rape

30	68	37.2	178	2 S25127	2S storage protein
31	68	37.2	180	2 S10018	napin (clone BngNA
32	67.5	36.9	37	2 A23617	conglutin delta-2
33	67	36.6	39	2 JQ2321	calmodulin antagonist
34	67	36.6	104	2 S48178	mablinin III - Yun
35	67	36.6	178	2 S25134	2S storage protein
36	66.5	36.3	100	2 S48180	mablinin IV - Yun
37	66.5	36.3	153	2 A33090	conglutin delta pr
38	66	36.1	33	2 P00801	calmodulin antagonist
39	66	36.1	133	1 NMRP1	napin I precursor
40	66	36.1	164	1 NMMU1	2S albumin I precu
41	66	36.1	186	2 A29802	napin precursor (9
42	65	35.5	39	2 P00802	calmodulin antagonist
43	64.5	35.2	166	1 NMMU4	2S albumin 4 precu
44	64	35.0	37	2 S16183	napin nIII - rape
45	63.5	34.7	164	1 NMMU3	2S albumin 3 precu

ALIGNMENTS

RESULT 1

RZCS

2S seed storage protein precursor - castor bean

N:Alternate names: 2S albumin precursor

C:Species: Ricinus communis (castor bean)

C>Date: 14-Nov-1993 #sequence;revision 08-Feb-1996 #text;change 18-Jun-1999

C/Accession: S11499; S11500; S11501; S11502; S27221; A01328; S27222

R:Irwin, S.D.; Lord, J.M.

Nucleic Acids Res. 18, 5890, 1990

A>Title: Nucleotide sequence of a Ricinus communis 2S albumin precursor gene.

A:Reference number: S11499; MUID:91016940

A/Accession: S11499

A:Molecule type: DNA

A:Residues: 1-258 <IRN>

A:Cross-references: EMBL:X54158; NID:Q21067; PIDN:CAA38097.1; PID:Q21068

A>Note: The authors translated the codon CTC for residue 14 as Phe, CCA for residue 7

R:Irwin, S.D.; Keen, J.N.; Findlay, J.B.C.; Lord, J.M.

Mol. Gen. Genet. 222, 400-408, 1990

A>Title: The Ricinus communis 2S albumin precursor: a single preproprotein may be pro

A:Reference number: S11500; MUID:91109729

A/Accession: S11500

A:Molecule type: mRNA

A:Residues: 1-13, 'F', 'L', 'S', 'T', '75-258 <IR2>

A:Experimental source: clone 1494

A/Accession: S11501

A:Molecule type: mRNA

A:Residues: 'M', '4', 'LS', '7-13', 'F', '15-21 <IRF>

A:Experimental source: clone 10a12

A/Accession: S27221

A:Molecule type: protein

A:Residues: 'X', '37', 'X', '39-45', 'X', '158-161', 'X', '163-174', 'X' <IRX>

J. Biol. Chem. 257, 14753-14759, 1982

A>Title: Amino acid sequence of small and large subunits of seed storage protein from

A:Reference number: A92357; MUID:83082772

A/Accession: A01328

A:Molecule type: protein

A:Residues: 157-190,194-221, 'Q', '223-225,230-233, 'N', '235-254, 'Q', '256-258 <SHA>

A>Note: 230-Set was also found

A>Note: There is considerable similarity between residues 181-231 of this protein and

R:Odani, S.; Koide, T.; Ono, T.; Ohnishi, K.

Biochem. J. 213, 543-545, 1983

A>Title: Structural relationship between barley (Hordeum vulgare) trypsin inhibitor a

A:Reference number: A90322; MUID:83308577

A:Contents: annotation

A>Note: this protein is homologous with trypsin inhibitor from barley

C:Complex: consists of two chains linked by two disulfide bonds involving Cys-162 and

C:Superfamily: 2S seed storage protein precursor

C:Keywords: pyroglutamic acid; seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:2-35/Domain: propeptide #status predicted <PRO>
F:36-72/Product: probable 2S seed storage protein small chain 2 #status experimental <SC>
F:87-156/Product: probable 2S seed storage protein small chain 2 #status predicted <LCHI>
F:157-190/Product: 2S seed storage protein small chain #status experimental <SML>
F:194-258/Product: 2S seed storage protein large chain #status experimental <LRG>
F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 100.0%; Score 183; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 4.4e-17;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PSQGRGQIQEQNLKRCQCEYIKQVSGGPRR 34
Db 157 PSQGRGQIQEQNLKRCQCEYIKQVSGGPRR 190

RESULT 2
S26636

napin n1b - rape

C:Species: Brassica napus (rape)

C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 04-Oct-1996

C:Accession: S26636

R:Monsalve, R.I.; Lopez-Otin, C.; Villalba, M.; Rodriguez, R.

FEBS Lett. 295, 207-210, 1991

A:Title: A new distinct group of 2 S albumins from rapeseed. Amino acid sequence of two

A:Reference number: S20350; MUID:92111741

A:Accession: S26636

A:Molecule type: protein

A:Residues: 1-106 <MON>

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: pyroglutamic acid; seed; storage protein

F:1-31/Product: napin small chain #status experimental <SMA>

F:32-106/Product: napin large chain #status experimental <LAR>

F:32/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen

Query Match 44.8%; Score 82; DB 2; Length 106;
Best Local Similarity 40.5%; Pred. NO. 0.00065;

Matches 15; Conservative 10; Mismatches 4; Indels 8; Gaps 1;

OY 4 QGCRGQIQEQNLKRCQCEYIKQVSG-----QGP 32
Db 3 QKCRFQEQEHLKACQGWIRQLAGSPFGSGPQGGP 39

RESULT 3

S20350

napin n1a - rape

C:Species: Brassica napus (rape)

C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 04-Oct-1996

C:Accession: S20350; S20351

R:Monsalve, R.I.; Lopez-Otin, C.; Villalba, M.; Rodriguez, R.

FEBS Lett. 295, 207-210, 1991

A:Title: A new distinct group of 2 S albumins from rapeseed. Amino acid sequence of two

A:Reference number: S20350; MUID:92111741

A:Accession: S20350

A:Molecule type: protein

A:Residues: 1-110 <MON>

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: pyroglutamic acid; seed; storage protein

F:1-31/Product: napin small chain #status experimental <SMA>

F:32-110/Product: napin large chain #status experimental <LAR>

F:32/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen

Query Match 43.7%; Score 80; DB 2; Length 110;
Best Local Similarity 41.2%; Pred. No. 0.0013;
Matches 14; Conservative 11; Mismatches 5; Indels 4; Gaps 1;

OY 4 QGCRGQIQEQNLKRCQCEYIKQVSG---OGPR 33

Db 3 QKCRFQEQEHLKACQGWIRQLAGSPFGSGPQ 36

RESULT 4

S14947

2S albumin - Brazil nut

C:Species: Bertholletia excelsa (Brazil nut)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000

C:Accession: S14947

R:Gander, E.S.; Holmstrom, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gr

Plant Mol. Biol. 16, 437-448, 1991

A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin

A:Reference number: S14946; MUID:91370890

A:Accession: S14947

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-154 <GAN>

A:Cross-references: EMBL:X54491; NID:917712; PIDN:CA38363.1; PID:917713

C:Genetics:

A:Introns: 67/3

C:Superfamily: wheat alpha-amylase inhibitor

Query Match 41.8%; Score 76.5; DB 2; Length 154;
Best Local Similarity 42.9%; Pred. No. 0.0052;
Matches 15; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

OY 2 SQGCRGQIQEQNLKRCQCEYIKQVSG---QGR 33
Db 43 SEQCRQEQERQDQNLKRCQCEYIKQVSGPYQNR 77

RESULT 5

T09878

albumin 2S storage protein precursor Mat5-A - upland cotton

C:Species: Gossypium hirsutum (upland cotton)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 29-Oct-1999

C:Accession: T09878

R:Galau, G.A.; Wang, H.Y.C.; Hughes, D.W.

submitted to the EMBL Data Library, January 1992

A:Description: Cotton Mat5-A (C164) gene and Mat5-D cDNAs encoding methionine-rich 2S

A:Reference number: Z16893

A:Accession: T09878

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-139 <GAL>

A:Cross-references: EMBL:M66213; NID:9167358; PID:9167359

C:Genetics:

A:Gene: Mat5-A

C:Keywords: storage protein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-139/Product: albumin 2S storage protein Mat5-A #status predicted <MAN>

Query Match 40.4%; Score 74; DB 2; Length 139;
Best Local Similarity 33.3%; Pred. No. 0.01;
Matches 10; Conservative 13; Mismatches 7; Indels 0; Gaps 0;

OY 2 SQGCRGQIQEQNLKRCQCEYIKQVSGG 31
Db 32 NRDSCEQIRKQAKLKHCKQKMEBELGEG 61

RESULT 6

B59346

seed storage protein Lec2SA1 small chain [imported] - tomato (fragment)

N:Alternate names: 2S albumin small subunit

C:Species: Lycopersicon esculentum (tomato)

C:Date: 31-Dec-2001 #sequence_revision 31-Dec-2001 #text_change 31-Dec-2001

C:Accession: B59346

R:Oguri, S.

submitted to the Protein Sequence Database, September 2000

A:Reference number: A59346
A:Accession: B59346
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-32 <OGU>
A:Experimental source: cv. cherry; seed
C:Complex: heterodimer of large (see PIR:A59346) and small chains, disulfide linked
C:Keywords: heterodimer; seed; storage protein

Query Match 39.8%; Score 73; DB 2; Length 32;
Best Local Similarity 42.3%; Pred. No. 0.0031;
Matches 11; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

OY 1 PSQGGCRGQIQEQNLROCCQERYIKQV 26
DB 1 PQTSCQGGPQQAQQLRSCQQLRQR 26

RESULT 7

A25802
2S seed storage protein small chain - Brazil nut

C:Species: Bertholletia excelsa (Brazil nut)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 19-May-2000
C:Accession: A25802

R:Ampe, C.; Van Damme, J.; de Castro, L.A.B.; Sampato, M.J.A.M.; Van Montagu, M.; Vandek
Eur. J. Biochem. 159, 597-604, 1986

A:Title: The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nu
A:Reference number: A91173; MUID:87004679

A:Accession: A25802
A:Molecule type: protein

A:Residues: 1-28 <AMP>
C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: pyrogutamic acid
F1/Modified site: pyroglutamic acid (Gln) #status experimental

Query Match 38.8%; Score 71; DB 2; Length 28;
Best Local Similarity 52.0%; Pred. No. 0.005;
Matches 13; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 3 QOGCRGQIQEQNLROCCQERYIKQV 27
DB 1 QOOCRFQMOQRLSHCRMYMRQM 25

RESULT 8

S35591
major allergen small chain - leaf mustard

C:Species: Brassica juncea (leaf mustard)
C:Date: 10-Dec-1993 #sequence_revision 27-Feb-1997 #text_change 07-May-1999
C:Accession: S35591

R:Monsalve, R.I.; Gonzalez de la Pena, M.A.; Menendez-Arias, L.; Lopez-Otin, C.; Villalal
Biochem. J. 293, 625-632, 1993

A:Title: Characterization of a new oriental-mustard (Brassica juncea) allergen, Bra j IF
A:Reference number: S35591; MUID:93356721

A:Accession: S35591
A:Status: preliminary

A:Molecule type: protein
A:Residues: 1-37 <MON>

C:Superfamily: wheat alpha-amylase inhibitor

Query Match 38.8%; Score 71; DB 2; Length 37;
Best Local Similarity 44.8%; Pred. No. 0.0066;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLROCCQERYIKQV--SGQGP 32
DB 9 CRKEFQQAQHLRACQQLHKKQAMQSGSGP 37

RESULT 9

S01791

allergen Sin a I small chain - white mustard

C:Species: Sinapis alba (white mustard)
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 25-Oct-1996

C:Accession: S01791
R:Menendez-Arias, L.; Moneo, I.; Dominguez, J.; Rodriguez, R.

Eur. J. Biochem. 177, 159-166, 1988

A:Title: Primary structure of the major allergen of yellow mustard (Sinapis alba L.)
A:Reference number: S01791; MUID:89030681

A:Accession: S01791
A:Molecule type: protein

A:Residues: 1-39 <MEN>
A:Note: 6-Gly was also found

C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed

Query Match 38.8%; Score 71; DB 2; Length 39;
Best Local Similarity 44.8%; Pred. No. 0.007;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLROCCQERYIKQV--SGQGP 32
DB 10 CRKEFQQAQHLRACQQLHKKQAMQSGSGP 38

RESULT 10

S65447
allergen Sin a I - white mustard

N:Alternate names: allergen Sin a I large chain
C:Species: Sinapis alba (white mustard)

C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S65447; S01792

R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996

A:Title: Expression in Escherichia coli of Sin a I, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251

A:Accession: S65447
A:Molecule type: protein

A:Residues: 1-127 <GON>
A:Experimental source: seed

R:Menendez-Arias, L.; Moneo, I.; Dominguez, J.; Rodriguez, R.
Eur. J. Biochem. 177, 159-166, 1988

A:Title: Primary structure of the major allergen of yellow mustard (Sinapis alba L.)
A:Reference number: S01791; MUID:89030681

A:Accession: S01792
A:Molecule type: protein

A:Residues: 40-127 <MEN>
C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 127;
Best Local Similarity 44.8%; Pred. No. 0.023;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLROCCQERYIKQV--SGQGP 32
DB 10 CRKEFQQAQHLRACQQLHKKQAMQSGSGP 38

RESULT 11

PC1246
Sin a I allergen I precursor - white mustard (fragment)

C:Species: Sinapis alba (white mustard)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: PC1246

R:de la Pena, M.A.G.; Villalba, M.; Garcia-Lopez, J.L.; Rodriguez, R.
Biochem. Biophys. Res. Commun. 190, 648-653, 1993

A:Title: Cloning and expression of the major allergen from yellow mustard seeds, Sin
A:Reference number: PC1246; MUID:93151856

A:Accession: PC1246
A:Molecule type: DNA

A:Residues: 1-145 <DLP>
A:Experimental source: seed
A>Note: the authors translated the codon GAG for residue 45 as Gln
A>Note: the species is designated as "yellow mustard"
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein
F:1-39/Domain: signal sequence <status predicted <SIG>
F:40-54/Domain: propeptide <status predicted <PRO>
F:55-145/Product: sin a I allergen 1 (fragment) <status predicted <MAT>

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQDNLRCQCEYIKQV--SGGCP 32
DB 10 CRKEFOAQHLRACQOWLHKQAMQSGSGP 38

RESULT 12

565478
allergen Sin a I (clone SIN5) - white mustard (fragment)
C:Species: Sinapis alba (white mustard)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S65478
R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996
A:Title: Expression in Escherichia coli of sin a 1, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251
A:Accession: S65478
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-145 <GON>
A:Cross-references: EMBL:X91798; NID:g1009441; PIDN:CAA62908.1; PID:g1009442
A:Experimental source: seed
C:Genetics:
A:Gene: sin5
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQDNLRCQCEYIKQV--SGGCP 32
DB 10 CRKEFOAQHLRACQOWLHKQAMQSGSGP 38

RESULT 13

565481
allergen Sin a I (clone SIN3) - white mustard (fragment)
C:Species: Sinapis alba (white mustard)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S65481
R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996
A:Title: Expression in Escherichia coli of sin a 1, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251
A:Accession: S65481
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-145 <GON>
A:Cross-references: EMBL:X91801; NID:g1009437; PIDN:CAA62911.1; PID:g1009438
A:Experimental source: seed
C:Genetics:
A:Gene: sin3
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQDNLRCQCEYIKQV--SGGCP 32
DB 10 CRKEFOAQHLRACQOWLHKQAMQSGSGP 38

RESULT 14

565482
allergen Sin a I (clone SIN4) - white mustard (fragment)
C:Species: Sinapis alba (white mustard)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S65482
R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996
A:Title: Expression in Escherichia coli of sin a 1, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251
A:Accession: S65482
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-145 <GON>
A:Cross-references: EMBL:X91802; NID:g1009439; PIDN:CAA62912.1; PID:g1009440
A:Experimental source: seed
C:Genetics:
A:Gene: sin4
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQDNLRCQCEYIKQV--SGGCP 32
DB 10 CRKEFOAQHLRACQOWLHKQAMQSGSGP 38

RESULT 15

565480
allergen Sin a I (clone SIN2) - white mustard (fragment)
C:Species: Sinapis alba (white mustard)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S65480
R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996
A:Title: Expression in Escherichia coli of sin a 1, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251
A:Accession: S65480
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-145 <GON>
A:Cross-references: EMBL:X91800; NID:g1009435; PIDN:CAA62910.1; PID:g1009436
A:Experimental source: seed
C:Genetics:
A:Gene: sin2
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQDNLRCQCEYIKQV--SGGCP 32
DB 10 CRKEFOAQHLRACQOWLHKQAMQSGSGP 38

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:46:06 ; Search time 22.97 Seconds
(without alignments)
57.312 Million cell updates/sec

Title: US-09-913-351-4
Perfect score: 183
Sequence: 1 PSQCGCGQIOEQONLRQCCEYIKQGVSGGPRR 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	183	100.0	258	1 2SS_RICCO	P01089 ricinus com
2	80	43.7	110	1 2SS1_BRANA	P24565 brassica na
3	72	39.3	129	1 ALL1_BRAJU	P80207 brassica ju
4	71	38.8	127	1 ALL1_SINAI	P15322 sinapis alu
5	71	38.8	130	1 ITRY_SINAR	P38057 sinapis arv
6	70	38.3	146	1 2SS_BEREX	P04403 bertholleti
7	69	37.7	104	1 2SS1_CAPWA	P80351 caparis ma
8	69	37.2	178	1 2SS2_BRANA	P01090 brassica na
9	68	37.2	178	1 2SS2_BRANA	P27740 brassica na
10	68	37.2	180	1 2SS3_BRANA	P17333 brassica na
11	67.5	36.9	37	1 CGS3_LUPAN	P09930 lupinus ang
12	67	36.6	104	1 2SS3_CAPWA	P80352 caparis ma
13	66.5	36.3	100	1 2SS4_CAPWA	P80353 caparis ma
14	66	36.1	133	1 2SS1_BRANA	P01091 brassica na
15	66	36.1	164	1 2SS1_ARATH	P15457 arabidopsis
16	66	36.1	186	1 2SS2_BRANA	P09893 brassica na
17	64.5	35.2	166	1 2SS4_ARATH	P15460 arabidopsis
18	63.5	34.7	164	1 2SS3_ARATH	P15458 arabidopsis
19	62.5	34.2	170	1 2SS2_ARATH	P15459 arabidopsis
20	58.5	32.0	158	1 2SS1_MOUSE	P19594 glycine max
21	57	31.1	182	1 INB_MOUSE	P19595 glycine max
22	56.5	30.9	705	1 FXPI_MOUSE	P84866 mus musculu
23	56	30.6	525	1 VCL_THRCC	O43358 theobroma c
24	55.5	30.3	304	1 GDB1_WHEAT	P04722 triticum ae
25	55	30.1	904	1 AAC4_CHICK	Q80734 gallus gall
26	54	29.5	313	1 GDAT_WHEAT	P04727 triticum ae
27	53.5	29.2	295	1 GLTC_WHEAT	P16135 triticum ae
28	53	29.0	144	1 IAAE_HORVU	P01086 hordeum vul
29	53	29.0	4349	1 DYHC_FUSOO	P78716 fusarium so
30	52.5	28.7	1060	1 FEH1_DROME	P28166 drosophila
31	52	28.4	155	1 2SS2_CAPWA	P20233 caparis ma
32	51	28.4	290	1 RAS1_CANAL	Q94890 drosophila
33	51	27.9	482	1 NF31_NABFO	P42661 naegleria f

34	51	27.9	905	1 SNF5_YEAST	P18480 saccharomyc
35	51	27.9	2038	1 FSH_DROME	P13709 drosophila
36	51	27.9	3828	1 TRX_DROVI	O24742 drosophila
37	50.5	27.6	289	1 HOG3_HORVU	P80198 hordeum vul
38	50.5	27.6	295	1 2SS5_HELAN	P15461 heliantus
39	50.5	27.6	648	1 GLT0_WHEAT	P10387 triticum ae
40	50.5	27.6	660	1 GLT3_WHEAT	P08488 triticum ae
41	50.5	27.6	1477	1 HRT7_HYDAT	O25197 hydra atten
42	50.5	27.6	3703	1 ABF1_HUMAN	O15911 homo sapien
43	50	27.3	435	1 NPL6_YEAST	P32832 saccharomyc
44	50	27.3	544	1 H15_DROME	O94890 drosophila
45	50	27.3	873	1 RX_DROME	O9w2q1 drosophila

ALIGNMENTS

```

RESULT 1
ID 2SS_RICCO STANDARD: PRT: 258 AA.
AC P01089;1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S albumin precursor (Allergen Ric c 1).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm.
RX MEDLINE=91108729; PubMed=2274038;
RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor: a single preproprotein
RT may be processed into two different heterodimeric storage proteins.";
RL Mol. Gen. Genet. 222:400-408(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91016940; PubMed=2216785;
RA Irwin S.D., Lord J.M.;
RT "Nucleotide sequence of a Ricinus communis 2S albumin precursor
RT gene.";
RL Nucleic Acids Res. 18:5890-5890(1990).
RN [3]
RP SEQUENCE OF 157-190 AND 194-258.
RX MEDLINE=83082772; PubMed=7174664;
RA Sharief F.S., Li S.S.-L.;
RT "Amino acid sequence of small and large subunits of seed storage
RT protein from Ricinus communis.";
RL J. Biol. Chem. 257:14753-14759(1982).
RN [4]
RP SIMILARITY TO PROTEINASE INHIBITORS.
RX MEDLINE=83308577; PubMed=6615448;
RA Odani S., Koide T., Ono T., Ohnishi K.;
RT "Structural relationship between barley (Hordeum vulgare) trypsin
RT inhibitor and castor-bean (Ricinus communis) storage protein.";
RL Biochem. J. 213:543-545(1983).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY 2 DISULFIDE BONDS.
CC -1- PTM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE
CC CHAINS INVOLVE CYS-162 AND CYS-175.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```


DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Allergen Sin a 1, small and large chains (Sin a 1).
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=89030681; PubMed=3181153;
RA Menendez-Arias L., Moneo I., Dominguez J., Rodriguez R.;
RT "Primary structure of the major allergen of yellow mustard (Sinapis
alba L.) seed, Sin a 1";
RL But. J. Biochem. 177:159-166(1988).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- DISEASE: ALLERGIC DISEASE, CABBAGE ALLERGY.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR PIR: S01792; S01792.
DR PIR: S01791; S01791.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR ProDom: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Allergen; Seed storage protein.
FT CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 127 LARGE CHAIN.
FT VARIANT 6 6 R->G.
SQ SEQUENCE 127 AA; 14180 MW; 4CD920284F04EEFD CRC64;

Query Match 38.8%; Score 71; DB 1; Length 127;
Best Local Similarity 44.8%; Pred. No. 0.014;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;
OY 6 CRGQIQEONLRCQCEYIKQV--SGGCP 32
DB 10 CRKEFOAQOHLRACQOQLHKQAMSGSGP 38
RESULT 5
ID ITTRY_SINAR STANDARD; PRT; 130 AA.
AC P38057;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin inhibitor (TISA).
OS Sinapis avensis (Charlock).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=29728;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=94350545; PubMed=8070965;
RA Svendsen I.B., Nicolova D., Goshvay I., Genov N.;
RT "Primary structure, spectroscopic and inhibitory properties of a two-
chain trypsin inhibitor from the seeds of charlock (Sinapis avensis
L.), a member of the napin protein family";
RL Int. J. Pept. Protein Res. 43:425-430(1994).
CC -1- FUNCTION: INHIBITS TRYPSIN WITH A KI OF 7 X 10(-6) M.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.

DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR ProDom: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Multigene family; Polymorphism.
FT CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 130 LARGE CHAIN.
FT VARIANT 32 32 R->M.
FT VARIANT 53 57 MISSING (IN FORM II).
FT VARIANT 73 73 A->S.
FT VARIANT 77 77 K->R.
FT VARIANT 81 81 Q->R.
FT VARIANT 87 87 H->Q.
FT VARIANT 89 89 Q->H.
FT VARIANT 91 91 G->Q.
FT VARIANT 97 97 E->M.
FT VARIANT 98 98 I->V.
FT VARIANT 99 98 R->S.
FT VARIANT 106 106 T->K.
FT VARIANT 123 123 N->Q.
FT VARIANT 124 124 K->G.
FT VARIANT 126 126 M->V.
SQ SEQUENCE 130 AA; 14682 MW; EC02EA826D180DF2 CRC64;

Query Match 38.8%; Score 71; DB 1; Length 130;
Best Local Similarity 44.8%; Pred. No. 0.014;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEONLRCQCEYIKQV--SGGCP 32
DB 10 CRKEFOAQOHLRACQOQLHKQAMSGSGP 38

RESULT 6
ID 2SS_BEREX STANDARD; PRT; 146 AA.
AC P04403; P04402;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S sulfur-rich seed storage protein precursor (Allergen Ber e 1).
GN BE2S1 AND BE2S2.
OS Bertholletia excelsa (Brazil nut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Lecythidaceae; Bertholletia.
OX NCBI_TaxID=3645;
RN [1]
RP SEQUENCE FROM N.A.
RA Alendach S.B., Pearson K.W., Leung F.W., Sun S.S.M.;
RT "Cloning and sequence analysis of a cDNA encoding a Brazil nut protein
exceptionally rich in methionine";
RL Plant Mol. Biol. 8:239-250(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Baasener R.;
RN Submitted (DEC-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91370890; PubMed=1840683;
RA Gander E.S., Holmstroem K.O., de Paiva G.R., de Castro L.A.B.,
RA Carneiro M., Grossi de Sa M.F.;
RT "Isolation, characterization and expression of a gene coding for a 2S
albumin from Bertholletia excelsa (Brazil nut)";
RL Plant Mol. Biol. 16:437-448(1991).
RN [4]
RP SEQUENCE OF 37-64 AND 70-142.
RX MEDLINE=87004679; PubMed=3758080;
RA Ampe C., van Damme J., de Castro L.A.B., Sampaio M.J.A.M.,
RA van Montagu M., Vandekerckhove J.;

```
RT      "The amino-acid sequence of the 2S sulphur-rich proteins from seeds  
FT    of Brazil nut (Berthollletia excelsa H.B.K.).";  
RL     Eur. J Biochem. 159:597-604(1986).  
CC     - I- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.  
CC     - I- SUBUNIT: THE MAJURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN  
CC     LINKED BY DISULFIDE BONDS.  
CC     - I- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.  
-----  
CC     This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC     between the Swiss Institute of Bioinformatics and the EMBL outpost at  
CC     the European Bioinformatics Institute. There are no restrictions on its  
CC     use by non-profit institutions as long as its content is in no way  
CC     modified and this statement is not removed. Usage by and for commercial  
CC     entities requires a license agreement (see http://www.isb-sdb.ch/announce/  
CC     or send an email to licenseseib@sdb.ch).  
-----  
DR     EMBL; M17146; AAA3010.1;   .;  
DR     EMBL; X57027; CAA40343.1;   .;  
DR     EMBL; X57028; CAA40344.1;   .;  
DR     EMBL; X54490; CAA38362.1;   .;  
DR     EMBL; X54491; CAA38363.1; ALT_SEQ.  
DR     EMBL; AI3818; CAA01131.1;   .;  
DR     PIR; S06252; S06252.  
DR     PIR; A25802; A25802.  
DR     PIR; B25802; B25802.  
DR     PIR; S14946; S14946.  
DR     PIR; S14479; S14479.  
DR     PIR; S21640; S21640.  
DR     InterPro: IPRO003612; AAI.  
DR     InterPro: IPRO01768; Cereal_ryp_amyl_inh.  
DR     Pfam; PF00234; tryp_alpha_amyl_1.  
DR     SMART; SM00499; AA1; 1.  
KW     Seed storage protein; Signal; Allergen.  
FT     SIGNAL          1       22  
FT     PROPEP         23       36  
FT     CHAIN           37       64        SMALL CHAIN.  
FT     PROPEP         65       69  
FT     CHAIN           70      142        LARGE CHAIN 1B.  
FT     PROPEP        143      146  
FT     MOD_RES        37       37  
FT     VARIANT        91       91        PYRROLIDONE CARBOXYLIC ACID.  
FT     CONFLICT       38       39        EE->QQ (IN REF. 4) .  
FT     CONFLICT       122      122        L->M (IN REF. 4) .  
FT     CONFLICT       126      126        I->L (IN REF. 4) .  
SQ     SEQUENCE        146 AA; 16911 MM; AVDF78FD766410D CRC64;  
  
Query Match                      38.3%; Score 70; DB 1; Length 146;  
Best Local Similarity            40.5%; Pred. No. 0.022;  
Matches 15; Conservative 7; Mismatches 11; Indels 4; Gaps 1;  
  
QY      2 SQQGCGGQIOEQNLROCCEVYIKGVSGG----PRR 34  
       :|::||:|| || | ::||::|:  
Db      36 NOEBREEMQMORQKMSHCRRMYRQGMESSPYTMR 72  
  
RESULT 7  
AC     2SS1_CAPMA STANDARD; FRT: 104 AA.  
DT     01-FEB-1995 (Rel. 31, Created)  
DT     01-FEB-1995 (Rel. 31, Last sequence update)  
DT     16-OCT-2001 (Rel. 40, Last annotation update)  
DE     Mablinin I-L, A and B chains (MAB I) (Sweet protein).  
OS     Capparis masaikei (Mabiniang).  
OC     Euaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC     Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC     eurossids II; Brassicales; Brassicaceae; Caparis.  
OX     NCBI_TaxId=13395;  
RN     [1]  
RP     SEQUENCE.  
RC     TISSUE=Seed;  
RX     MEDLINE=94333405; PubMed=8055976;
```

RA Nirasawa., Nishino T., Katalaura M., Uesugi S., Hu Z., Kurihara Y.,
 RT "Structures of heat-stable and unstable homologues of the sweet
 RT protein mabinlin. The difference in the heat stability is due to
 RT replacement of a single amino acid residue.";
 RL Eur. J. Biochem. 223:989-995(1994).
 CC -1- FUNCTION: 2S SEED STORAGE PROTEIN HAVING SWEETNESS-INDUCING
 CC ACTIVITY. THIS FORM IS NOT HEAT STABLE.
 CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
 CC DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF01631; Seedstore_2S; 1.
 DR PRINTS: PR00496; NAPIN.
 DR ProDom: PD002498; Napin; 1.
 DR SMART: SM00499; AAI; 1.
 KW Seed storage protein; Albumin; Sweet-taste.
 FT CHAIN 1 32 MABINLIN I, A CHAIN.
 FT NON CONS 32 33
 FT CHAIN 33 104 MABINLIN I, B CHAIN.
 FT DISULFID 4 53
 FT DISULFID 17 42
 FT DISULFID 43 91
 FT DISULFID 55 99
 SQ SEQUENCE 104 AA; 12284 MW; 156AA4AB50D5EE60 CRC64;
 Query Match 37.7%; Score 69; DB 1; Length 104;
 Best Local Similarity 46.2%; Pred. No. 0.021;
 Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 QY 6 CRGQIQEQNLRQCCQETIKQVSGG 31
 ||| | : ||| | | : : |
 Db 4 CRQFQHQHRLRACQRYIRRAQRCG 29
 RESULT 8
 2SS2_BRANA STANDARD; PRT: 178 AA.
 ID 2SS2_BRANA
 AC P01090;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE Napin 2 precursor (1.7S seed storage protein).
 OS Brassica napus (Rapeseed).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eustosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87308224; PubMed=3624251;
 RA Josefsson L.-G., Lennan M., Ericsson M.L., Rask L.:
 RT "Structure of a gene encoding the 1.7 S storage protein, napin, from
 RT Brassica napus.";
 RL J. Biol. Chem. 262:12196-12201(1987).
 RN [2]
 RP REVISIONS.
 RA Josefsson L.-G.:
 RL Submitted (JUL-1987) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8703365; PubMed=3771543;
 RA Ericsson M.L., Reedin J., Lennan M., Glimelius K., Josefsson L.-G.,
 RA Rask L.:
 RT "Structure of the rapeseed 1.7 S storage protein, napin, and its
 RT precursor.";
 RL J. Biol. Chem. 261:14576-14581(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, TOWER;
 RX MEDLINE=84113267; PubMed=6689334;
 RA Crouch M.L., Tenbarghe K.M., Simon A.E., Ferl R.:

RT "cDNA clones for Brassica napus seed storage proteins: evidence from
 RT nucleotide sequence analysis that both subunits of napin are cleaved
 RT from a precursor polypeptide.";
 RL J. Mol. Appl. Genet. 2:273-283(1983).
 CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
 CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
 CC ITS MATURATION.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY DISULFIDE BONDS.
 CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; K01545; AAA33006.1; -;
 CC EMBL; J02586; AAA3297.1; -;
 CC EMBL; J02798; AAA87348.1; -;
 CC PIR; A01329; NMRP2.
 CC PIR; A29801; A29801.
 CC PIR; A25997; A25997.
 CC InterPro; IPR003612; AAI.
 CC InterPro; IPR001768; Cereal-trypan-amy1_inh.
 CC InterPro; IPR000617; Napin.
 CC Pfam; PF00234; trypan-amy1; 1.
 CC PRINTS; PR00496; Napin.
 CC ProDom; PD002498; Napin; 1.
 CC SMART; SM00499; AAI; 1.
 CC Seed storage protein; Signal; Multigene family.
 FT SIGNAL 1 21
 FT PROPEP 22 38
 FT CHAIN 39 74 SMALL CHAIN.
 FT PROPEP 75 94
 FT CHAIN 95 175 LARGE CHAIN.
 FT CONFLICT 37 37 D -> N (IN REF. 4).
 FT CONFLICT 76 76 S -> N (IN REF. 4).
 SQ SEQUENCE 178 AA; 20104 MW; 734E561971B539FF CRC64;

Query Match 37.7%; Score 69; DB 1; Length 178;
 Best Local Similarity 44.8%; Pred. No. 0.035;
 Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGOIOEONLRQCOEYIKQV--SGGPG 32
 ID 2SSB_BRANA STANDARD; PRT; 178 AA.
 AC P27740.
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Napin B precursor (1.7S seed storage protein).
 GN NABP.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 CX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. SVALOFS KARAT 20516-K;
 RX MEDLINE=91231016; PubMed=2029903;
 RA Ericson M.L., Muren E., Gustavsson H.O., Josefsson L.G., Raak L.;
 RT "Analysis of the promoter region of napin genes from Brassica napus

RT demonstrates binding of nuclear protein in vitro to a conserved
 RT sequence motif.";
 RL Eur. J. Biochem. 197:741-746(1991).
 CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
 CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
 CC ITS MATURATION.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY DISULFIDE BONDS.
 CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; X58142; CAA41150.1; -;
 CC PIR; S15382; S15382.
 CC InterPro; IPR003612; AAI.
 CC InterPro; IPR001768; Cereal-trypan-amy1_inh.
 CC InterPro; IPR000617; Napin.
 CC Pfam; PF00234; trypan-amy1; 1.
 CC PRINTS; PR00496; Napin.
 CC ProDom; PD002498; Napin; 1.
 CC SMART; SM00499; AAI; 1.
 CC Seed storage protein; Signal; Multigene family; Embryo.
 FT SIGNAL 1 21
 FT PROPEP 22 38
 FT CHAIN 39 74 SMALL CHAIN (BY SIMILARITY).
 FT PROPEP 75 94 BY SIMILARITY.
 FT CHAIN 95 178 LARGE CHAIN (BY SIMILARITY).
 SQ SEQUENCE 178 AA; 20114 MW; 96CE0ADB7CD966E9 CRC64;

Query Match 37.2%; Score 68; DB 1; Length 178;
 Best Local Similarity 41.4%; Pred. No. 0.048;
 Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGOIOEONLRQCOEYIKQV--SGGPG 32
 ID 2SS3_BRANA STANDARD; PRT; 180 AA.
 AC P17333.
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Napin precursor (1.7S seed storage protein).
 GN NABP.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 CX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. WESTAR; TISSUE=Leaf;
 RX MEDLINE=91346554; PubMed=2102844;
 RA Baszczyński C.L., Fallis L.;
 RT "Isolation and nucleotide sequence of a genomic clone encoding a new
 RT Brassica napus napin gene.";
 RL Plant Mol. Biol. 14:633-635(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. TOPAS;
 RX MEDLINE=95161697; PubMed=7858212;
 RA Boutilier K.A., Gines M.J., Demoor J.M., Huang B.,

RA Baszczynski C.L., Iyer V.N., Miki B.L.;
 RT "Expression of the Binnap subfamily of napin genes coincides with the
 RT induction of Brassica microspore embryogenesis.";
 RL Plant Mol. Biol. 26:1711-1723(1994).
 CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
 CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
 CC ITS MATURATION.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY DISULFIDE BONDS.
 CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X17542; CAA35580.1; -;
 DR EMBL: U04945; AAA81909.1; -;
 DR PIR: S10018; S10018
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1-trypanin.
 DR Pfam: PF00234; trypanin_1.
 DR PRINTS: PR00496; NAPIN.
 DR ProDom: PD002498; Napin; 1.
 DR SMART: SM00499; AAI; 1.
 DR Seed storage protein; Signal; Multigene family.
 KW SIGNAL 1 21
 FT PROPEP 22 38 SMALL CHAIN.
 FT CHAIN 39 74
 FT PROPEP 75 94
 FT CHAIN 95 180 LARGE CHAIN.
 SQ SEQUENCE 180 AA; 20318 MW; 6F3883CBED55FB26 CRC64;
 Query Match 37.2%; Score 68; DB 1; Length 180;
 Best Local Similarity 41.4%; Pred. No. 0.048;
 Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 1;
 QY 6 CREGIOEQNLRCOCQEFYKQGV--SGQCP 32
 ID CG2S_LUPAN STANDARD; PRT; 37 AA.
 AC P09930;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Conglutin delta-2 small chain.
 OS Lupinus angustifolius (Narrow-leaved blue lupine).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
 OX NCBI_TaxID=3871;
 RN [1]
 RP SEQUENCE.
 RC STRAIN-CV. WHITE;
 RA Lilley G.G., Inglis A.S.;
 RT "Amino acid sequence of conglutin delta, a sulfur-rich seed protein
 RT of Lupinus angustifolius L. Sequence homology with the C-III alpha-
 RT amylase inhibitor from wheat.";
 RL FEBS Lett. 195:235-241(1986).
 CC -1- SUBUNIT: DIMER OF A SMALL CHAIN AND A LARGE CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 DR PIR: A23617; A23617.

KW Seed.
 FT DISULFID 8 8 INTERCHAIN (WITH C-29 OF LARGE CHAIN).
 FT DISULFID 20 20 INTERCHAIN (WITH C-17 OR C-18 OF LARGE
 FT DOMAIN 29 37 CHAIN).
 FT SEQUENCE 37 AA; 4598 MW; 89784D55A5A193A CRC64;
 Query Match 36.9%; Score 67.5; DB 1; Length 37;
 Best Local Similarity 44.8%; Pred. No. 0.011;
 Matches 13; Conservative 7; Mismatches 8; Indels 1; Gaps 1;
 QY 2 SQGCRGQIQEQNLRCOCQEFYKQVSGQ 30
 ID 4 SEQSCRKRLQ-QVNLRCHEMHIDRIQQQ 31
 Db 4
 RESULT 12
 2SS3_CAPMA STANDARD; PRT; 104 AA.
 ID 2SS3_CAPMA
 AC P80352;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mabinlin III, A and B chains (MAB III) (Sweet protein).
 OS Capparis masakal (Mabinlan).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Capparis.
 OX NCBI_TaxID=13395;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Seed;
 RX MEDLINE=94333405; PubMed=8055976;
 RA Nirasawa S., Nishino T., Katahira M., Uesugi S., Hu Z., Kurihara Y.;
 RT "Structures of heat-stable and unstable homologues of the sweet
 RT protein mabinlin. The difference in the heat stability is due to
 RT replacement of a single amino acid residue.";
 RL Eur. J. Biochem. 223:989-995(1994).
 CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
 CC INDUCING ACTIVITY.
 CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
 CC DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 CC InterPro: IPR003612; AAI.
 CC InterPro: IPR000617; Napin.
 CC Pfam: PF01631; Seedstore_2S; 1.
 CC PRINTS: PR00496; NAPIN.
 CC ProDom: PD002498; Napin; 1.
 DR SMART: SM00499; AAI; 1.
 DR SMART: SM00499; AAI; 1.
 KW Seed storage protein; Albumin; Sweet-taste.
 FT CHAIN 1 32 MABINLIN III, A CHAIN.
 FT NON_CONS 32 33 MABINLIN III, A CHAIN.
 FT CHAIN 33 104 MABINLIN III, B CHAIN.
 FT DISULFID 4 53
 FT DISULFID 17 42
 FT DISULFID 43 91
 FT DISULFID 55 99
 SQ SEQUENCE 104 AA; 12284 MW; 102EE3F5F24AD3D0 CRC64;
 Query Match 36.6%; Score 67; DB 1; Length 104;
 Best Local Similarity 42.3%; Pred. No. 0.037;
 Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 QY 6 CREGIOEQNLRCOCQEFYKQVSGQ 31
 ID 4 CREGIOEQNLRCOCQEFYKQVSGQ 29
 Db 4
 RESULT 13
 2SS4_CAPMA STANDARD; PRT; 100 AA.
 ID 2SS4_CAPMA

AC P80353;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin IV, A and B chains (MAB IV) (Sweet protein).
OS Capparis masatakal (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Capparidaceae;
OC NCBI_TaxID=13395;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RA MEDLINE=94333405; PubMed=8055976;
RA Nirasawa S., Nishino T., Katsuhira M., Uesugi S., Hu Z., Kurihara Y.;
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.";
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
CC INDUCING ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 28 MABINLIN IV, A CHAIN.
FT NON_CONS 28 29
FT CHAIN 29 100 MABINLIN IV, B CHAIN.
FT DISULFID 4 49 BY SIMILARITY.
FT DISULFID 17 38 BY SIMILARITY.
FT DISULFID 39 87 BY SIMILARITY.
FT DISULFID 51 95 BY SIMILARITY.
SQ SEQUENCE 100 AA; 11928 MW; 7B09673FAB7793CA CRC64;

Query Match 36.3%; Score 66.5; DB 1; Length 100;
Best Local Similarity 41.4%; Pred. No. 0.042;
Matches 12; Conservative 8; Mismatches 8; Indels 1; Gaps 1;

OY 6 CRGQIQEONLRQCOEYIKQVSGGPPR 34
||| : ||| ||| : : : :
DB 4 CRGFOQHHLKACQRYLRRRAO-RGEOR 31

RESULT 14
2S5L_BRANA STANDARD; PRT; 133 AA.
ID 2S5L_BRANA
AC P01091;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Napin I precursor (1.75 seed storage protein) (Fragment).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OC NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, TOWER;
RX MEDLINE=84113267; PubMed=6689334;
RA Crouch M.L., Tenberge K.M., Simon A.E., Ferl R.;
RT "cDNA clones for Brassica napus seed storage proteins: evidence from
RT nucleotide sequence analysis that both subunits of napin are cleaved
RT from a precursor polypeptide.";
RL J. Mol. Appl. Genet. 2:273-283(1983).
CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE

CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC ITS MATURATION.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- TISSUE SPECIFICITY: CORNEDONS AND THE AXIS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: K01544; AAA33005.1; -
DR PIR: A01330; NWRL1.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereals_1ryp_aml_1nh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; 1ryp_alpha_aml1; 1.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Multigene family.
FT CHAIN 1 1
FT NON_TER 1 1
FT CHAIN <1 30 SMALL CHAIN.
FT PROPEP 31 49
FT CHAIN 50 130
SQ SEQUENCE 133 AA; 15294 MW; 5FD5C7E3E22ACB CRC64;

Query Match 36.1%; Score 66; DB 1; Length 133;
Best Local Similarity 41.4%; Pred. No. 0.065;
Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEONLRQCOEYIKQV-SGQGP 32
||| : ||| ||| : : : :
DB 3 CRKEFOQHHLKACQRYLRRRAO-RGEOR 31

RESULT 15
2S5L_ARATH STANDARD; PRT; 164 AA.
ID 2S5L_ARATH
AC P15457;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S seed storage protein I precursor (2S albumin storage protein)
DE (NMWU2-2S albumin 1).
CN AT2S1 OR AT4G27140 OR T24A18.90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-73 AND 84-162.
RC STRAIN=CV, C24;
RA Kreibers E., Herdies L., de Clercq A., Seurinck J., Leemans J.,
RA van Damme J., Segura M., Cheysen G., van Montagu M.,
RA Vandekerckhove J.;
RT "Determination of the processing sites of an Arabidopsis 2S albumin
RT and characterization of the complete gene family.";
RL Plant Physiol. 87:859-866(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, C24;
RA Conceicao A.D.S., Kreibers E.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hehse J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Gymonprez B., Chuang Y.-J., Vandebussche F.,
RA Breken M., Welfjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzengger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzner E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moolman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
RA Benesler S., Hempel S., Feldpausch M., Lambert S., Van den Daele H.,
RA De Keyser A., Busschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen K.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., Mclay K., Mayes R.,
RA Petrett A., Rajendram M.A., Lyne M., Benes V., Neumann S.,
RA Borikova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Meiller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,
RA Neumann S., Argitjon A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Ougley F., Clabaud G., Mendenlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Glibbons T., Weber N., Vandepol M., Baigues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis S., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Meyers H.-W., Stocker S.,
RA Zaccaria P., Bervan M., Wilson R.K., de la Bastide W., Habermann K.,
RA Parnell L., Dedila N., Gnoj L., Schult K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Marx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spleth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.,
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana.";
RL Nature 402:769-777(1999).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- MISCELLANEOUS: THIS IS THE MOST ABUNDANT ISOFORM OF 2S ALBUMIN IN
CC ARABIDOPSIS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22032; AAA32743.1; -;
DR EMBL; Z24745; CAA80870.1; -;
DR EMBL; A13820; CAA01132.1; -;
DR EMBL; A1035680; CAB38844.1; -;
DR EMBL; A161566; CAB79369.1; -;
DR PIR; JA0161; NMMU1.
DR PIR; S34676; S34676.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_amy1_inh.
DR InterPro: IPR000617; Napin.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR PRINTS; PR00496; NAPIN.

DR ProDom: PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
KW Seed storage protein; Albumin; Signal; Multigene family.
FT SIGNAL 1 21
FT PROPEP 22 37
FT CHAIN 38 73
FT PROPEP 74 83
FT CHAIN 84 162
FT PROPEP 163 164
SQ SEQUENCE 164 AA; 19014 MW; 2BF28CB474D9832B CRC64;

Query Match 36.1%; Score 66; DB 1; Length 164;
Best Local Similarity 42.3%; Pred. No. 0.08;
Matches 11; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
OY 1 PSQCRCRGQIQEQONTLRQCQERYIKQ 26
DB 41 PKMRCKRFQKQKHTRACQQLMQLQQ 66

Search completed: September 9, 2002, 12:46:06
Job time: 308 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:45:37 : Search time 77.42 Seconds
(without alignments)
75.973 Million cell updates/sec

Title: US-09-913-351-4
Perfect score: 183
Sequence: 1 PSQGGCRGQIQEONLRQCCQRYIKQVSGGQPRR 34

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_oranelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	183	100.0	35	10	Q9S873	Q9S873 ricinus com
2	88	48.1	139	10	P93198	P93198 Juglans reg
3	74	40.4	139	10	Q39795	Q39795 gossypium h
4	72	39.3	125	10	P80208	P80208 brassica na
5	71	38.8	40	10	Q9S9F2	Q9S9F2 brassica na
6	71	38.8	145	10	Q41196	Q41196 sinapis alb
7	71	38.8	145	10	Q41281	Q41281 sinapis alb
8	71	38.8	145	10	Q41277	Q41277 sinapis alb
9	71	38.8	145	10	Q41278	Q41278 sinapis alb
10	71	38.8	145	10	Q41279	Q41279 sinapis alb
11	71	38.8	145	10	Q41280	Q41280 sinapis alb
12	71	38.8	152	10	Q41168	Q41168 rapanus sa
13	71	38.8	174	10	Q41169	Q41169 rapanus sa
14	71	38.8	180	10	Q42469	Q42469 brassica na
15	70	38.3	41	10	Q9S9F1	Q9S9F1 brassica na
16	70	38.3	139	10	Q39787	Q39787 gossypium h

17	70	38.3	146	10	Q9LR02	Q9LR02 berthollet
18	69	37.7	178	10	Q42490	Q42490 brassica ol
19	69	37.7	178	10	Q39344	Q39344 brassica na
20	68	37.2	178	10	Q42473	Q42473 brassica ca
21	68	37.2	184	10	Q42444	Q42444 brassica ca
22	67	36.6	178	10	Q42491	Q42491 brassica na
23	66.5	36.3	153	10	Q99235	Q99235 lupinus ang
24	66	36.1	39	10	Q9S802	Q9S802 rapanus ang
25	66	36.1	39	10	Q9S802	Q9S802 rapanus ang
26	65	35.5	178	10	Q9S339	Q9S339 brassica ca
27	65	35.5	39	10	Q9S801	Q9S801 rapanus sa
28	63	34.4	39	10	Q9S803	Q9S803 rapanus sa
29	62	33.9	155	10	Q41167	Q41167 rapanus sa
30	61	33.3	178	10	Q42413	Q42413 brassica ju
31	61	33.3	323	10	Q39928	Q39928 helianthus
32	61	33.3	528	10	Q80838	Q80838 arabidopsis
33	61	33.3	528	10	Q948V4	Q948V4 arabidopsis
34	60	32.8	882	11	Q9D4H4	Q9D4H4 mus musculu
35	59.5	32.5	41	10	Q9S874	Q9S874 ricinus com
36	59	32.2	30	10	Q9S8Y8	Q9S8Y8 rapanus sa
37	58	31.7	295	10	Q946W1	Q946W1 zea mays (m
38	58	31.7	378	5	Q27383	Q27383 caenorhabd
39	57.5	31.4	386	5	Q9W353	Q9W353 drosophila
40	57.5	31.4	618	5	Q9XVY8	Q9XVY8 caenorhabd
41	57	31.1	635	5	Q18094	Q18094 caenorhabd
42	57	31.1	388	5	Q16500	Q16500 caenorhabd
43	57	31.1	388	5	Q16501	Q16501 caenorhabd
44	57	31.1	438	5	Q16502	Q16502 caenorhabd
45	56.5	30.9	445	5	Q16511	Q16511 caenorhabd
			141	10	Q39649	Q39649 cucurbita s

ALIGNMENTS

RESULT 1
ID Q9S873 PRELIMINARY; PRT; 35 AA.
AC Q9S873;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE 4 KDA NAPIN-LIKE PROTEIN SMALL CHAIN (FRAGMENT).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_Taxid=3988;
RN [1]
RP SEQUENCE.
RX MEDLINE-97135090: PubMed-8980648:
RA Neumann G.M., Condron R., Polya G.M.;
RT Purification and sequencing of napin-like protein small and large
RT chains from Momordica charantia and Ricinus communis seeds and
RT determination of sites phosphorylated by plant Ca(2+)-dependent
RT protein kinase.";
RL Biochim. Biophys. Acta 1298:223-240(1996).
SQ SEQUENCE 35 AA; 4044 MW; 992B9BA8D6A68FD3 CRC64;

Query Match 100.0%; Score 183; DB 10; Length 35;
Best Local Similarity 100.0%; Pred. No. 4.8e-19;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSQGGCRGQIQEONLRQCCQRYIKQVSGGQPRR 34
DB 1 PSQGGCRGQIQEONLRQCCQRYIKQVSGGQPRR 34
P93198
ID P93198 PRELIMINARY; PRT; 139 AA.
AC P93198;
DT 01-MAY-1997 (TREMBLrel. 03, Created)

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 AC ALBUMIN SEED STORAGE PROTEIN (FRAGMENT).
 OS Juglans regia (English walnut).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Fagales; Juglandaceae; Juglans.
 OX NCBI_TaxID=51240;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. SUNLAND;
 RA Teuber S.S., Dandekar A.M., Peterson W.R.;
 RT "Juglans regia 2S albumin seed storage protein precursor.";
 RL Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U66866; ABA41308.1; -;
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_tryp_aml_1nh.
 DR InterPro: IPR004480; Glutelin.
 DR Pfam: PF00234; tryp_alpha_aml_1.
 DR PRINTS: PR00211; GLUTELIN.
 DR SMART: SM00499; AAI; 1.
 FT NON_TER 1
 SQ SEQUENCE 139 AA; 16373 MW; 02D0E55E67164F23 CRC64;

Query Match 48.1%; Score 88; DB 10; Length 139;
 Best Local Similarity 57.1%; Pred. No. 6.4e-05;
 Matches 16; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 4 CGCGQIQEQNLROCOEYIKQVSGG 31
 DB 37 ECGREGIQOQNLHNCQYLRQSGG 64

RESULT 3
 O39795 PRELIMINARY; PRT; 139 AA.
 AC O39795;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 2S ALBUMIN STORAGE PROTEIN PRECURSOR.
 GN Mat5-A.
 OS Gossypium hirsutum (Upland cotton).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Malvales; Malvaceae; Gossypium.
 OX NCBI_TaxID=3635;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COYYLEDON;
 RA Galau G.A., Wang H.Y.-C., Hughes D.W.;
 RT "Cotton Mat5-A (C164) gene and Mat5-D cDNAs encoding methionine-rich
 2S albumin storage proteins.";
 RL Submitted (JAN-1992) to the EMBL/Genbank/DBJ databases.
 DR EMBL: M86213; AAA33066.1; -;
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_tryp_aml_1nh.
 DR Pfam: PF00234; tryp_alpha_aml_1.
 DR SMART: SM00499; AAI; 1.
 KW Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 33 59 2S ALBUMIN STORAGE PROTEIN.
 FT CHAIN 64 139 2S ALBUMIN STORAGE PROTEIN.
 SQ SEQUENCE 139 AA; 15700 MW; 02ACE24FECE9FE90 CRC64;

Query Match 40.4%; Score 74; DB 10; Length 139;
 Best Local Similarity 33.3%; Pred. No. 0.0063;
 Matches 10; Conservative 13; Mismatches 7; Indels 0; Gaps 0;

OY 2 SOCGRGQIQEQNLROCOEYIKQVSGG 31
 DB 37 ECGREGIQOQNLHNCQYLRQSGG 64

DB 32 NRDSCEQOIRKQAHKHCQKMEELGGEG 61

RESULT 4
 P80208 PRELIMINARY; PRT; 125 AA.
 AC P80208;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE NAPIN BNIII (1.7S SEED STORAGE PROTEIN).
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=SEED;
 RX MEDLINE=97390373; PubMed=9249277;
 RA Monsalve R.I., Gonzalez de la Pena M.A., Lopez-otin C., Flandor A.,
 RA Fernandez C., Villalba M., Rodriguez R.;
 RT "Detection, isolation and complete amino acid sequence of an
 RT aroclallergenic protein from rapeseed flour.";
 RL Clin. Exp. Allergy 27:833-841(1997).
 CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
 CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
 CC ITS MATURATION.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY TWO DISULFIDE BONDS.
 CC -1- TISSUE SPECIFICITY: COYYLEDONS AND THE AXIS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_tryp_aml_1nh.
 DR InterPro: IPR00617; Napin.
 DR Pfam: PF00234; tryp_alpha_aml_1.
 DR PRINTS: PR00496; NAPIN.
 DR Prodom: PD002498; Napin; 1.
 DR SMART: SM00499; AAI; 1.
 KW Seed storage protein; Multigene family.
 FT CHAIN 1 37 SMALL CHAIN.
 FT CHAIN 38 125 LARGE CHAIN.
 SQ SEQUENCE 125 AA; 14035 MW; C4042A1BF07099E8 CRC64;

Query Match 39.3%; Score 72; DB 10; Length 125;
 Best Local Similarity 43.3%; Pred. No. 0.011;
 Matches 13; Conservative 8; Mismatches 7; Indels 2; Gaps 1;

OY 6 CGGQIQEQNLROCOEYIKQV--SGQGR 33
 DB 10 CKRFGQAHFLACQOWLHKQAMQSGSGPQ 39

RESULT 5
 O9S9F2 PRELIMINARY; PRT; 40 AA.
 AC O9S9F2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NAPIN SHORT CHAIN S3A-CALMODULIN ANTAGONIST/CALCIUM-DEPENDENT PROTEIN
 DE KINASE SUBSTRATE.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=96283790; PubMed=8679670;
 RA Neumann G.M., Condron R., Thomas I., Polya G.M.;
 RT "Purification and sequencing of multiple forms of Brassica napus seed


```

RESULT 9
Q41278
ID 041278 PRELIMINARY; PRT; 145 AA.
AC 041278:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALLERGEN SIN A 1.0105 (FRAGMENT).
GN SIN2.
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RA Gonzalez de la Peña M.A., Monsalve R.I., Batanero E., Villalba M.,
RA Rodríguez R.;
RT "Expression in Escherichia coli of the major allergen from mustard,
RT Sin a 1.";
RL J. Biol. Chem. 237:827-832(1996).
DR EMBL: X91800; CA62910.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_aml_1nh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; tryp_alpha_aml1; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 16418 MW; 7170F7B8AD78157 CRC64;

Query Match 38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLRCQCEYIKQV--SGQSP 32
DB 10 CRKEFQAQHLRACQQLHKQAMQSGSP 38

RESULT 10
Q41279
ID 041279 PRELIMINARY; PRT; 145 AA.
AC 041279:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALLERGEN SIN A 1.0106 (FRAGMENT).
GN SIN3.
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RA Gonzalez de la Peña M.A., Monsalve R.I., Batanero E., Villalba M.,
RA Rodríguez R.;
RT "Expression in Escherichia coli of the major allergen from mustard,
RT Sin a 1.";
RL J. Biol. Chem. 237:827-832(1996).
DR EMBL: X91801; CA62911.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_aml_1nh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; tryp_alpha_aml1; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 16462 MW; 16EB8E5CC2C07F2F CRC64;

```

```

DR SMART: SM00499; AAI; 1.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 16268 MW; 03158FE7D6CBD521 CRC64;

Query Match 38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLRCQCEYIKQV--SGQSP 32
DB 10 CRKEFQAQHLRACQQLHKQAMQSGSP 38

RESULT 11
Q41280
ID 041280 PRELIMINARY; PRT; 145 AA.
AC 041280:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALLERGEN SIN A 1.0107 (FRAGMENT).
GN SIN4.
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RA Gonzalez de la Peña M.A., Monsalve R.I., Batanero E., Villalba M.,
RA Rodríguez R.;
RT "Expression in Escherichia coli of the major allergen from mustard,
RT Sin a 1.";
RL J. Biol. Chem. 237:827-832(1996).
DR EMBL: X91802; CA62912.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_aml_1nh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; tryp_alpha_aml1; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 16462 MW; 16EB8E5CC2C07F2F CRC64;

Query Match 38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLRCQCEYIKQV--SGQSP 32
DB 10 CRKEFQAQHLRACQQLHKQAMQSGSP 38

RESULT 12
Q41168
ID 041168 PRELIMINARY; PRT; 152 AA.
AC 041168:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE STORAGE PROTEIN (FRAGMENT).
GN NAPIN.
OS Raphanus sativus (Radish).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Raphanus.
OX NCBI_TaxID=3726;

```


RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-IMMATURE SEED;
 RX MEDLINE-91216448; PubMed-2022325;
 RA Raynal M., Deligny D., Grellier F., Delseny M.;
 RT "Characterization and evolution of napin-encoding genes in radish and
 related crucifers.";
 RL Gene 99:77-86(1991).
 DR EMBL: M63842; AAA63471.1;
 DR InterPro: IPR003612; AAT.
 DR InterPro: IPR001768; Cereals-trypan-amylin.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; trypan-amylin.1.
 DR PRINTS: PR00496; NAPIN.
 DR SMART: SM00499; AAT; 1.
 FT NON_TER 1 1
 FT CHAIN 7 41
 FT CHAIN 61 150
 SQ SEQUENCE 152 AA; 17459 MW; 9885CACEA6C0755 CRC64;
 STORAGE PROTEIN SMALL SUBUNIT.

Query Match 38.8%; Score 71; DB 10; Length 152;
 Best Local Similarity 44.8%; Pred. No. 0.018;
 Matches 13; Conservative 7; Mismatches 2; Gaps 1;

OY 6 CRGQIQEONLRQCCQYIKQV--SGQGP 32
 DB 16 CRREFQAOHLRACQOMLHRQARQSGSGP 44

RESULT 13
 ID 041169 PRELIMINARY; PRT; 174 AA.
 AC 041169;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE STORAGE PROTEIN (FRAGMENT).
 OS NAPIN.
 GN Raphanus sativus (Radish).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Raphanus.
 OX NCBI_TaxID=3726;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-IMMATURE SEED;
 RX MEDLINE-91216448; PubMed-2022325;
 RA Raynal M., Deligny D., Grellier F., Delseny M.;
 RT "Characterization and evolution of napin-encoding genes in radish and
 related crucifers.";
 RL Gene 99:77-86(1991).
 DR EMBL: M63843; AAA63472.1;
 DR InterPro: IPR003612; AAT.
 DR InterPro: IPR001768; Cereals-trypan-amylin.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; trypan-amylin.1.
 DR PRINTS: PR00496; NAPIN.
 DR ProDom: PD002498; Napin.1.
 DR SMART: SM00499; AAT; 1.
 FT NON_TER 1 1
 FT CHAIN 7 41
 FT CHAIN 61 150
 SQ SEQUENCE 174 AA; 19928 MW; 966F0CF7B19A/FPS CRC64;
 STORAGE PROTEIN SMALL SUBUNIT.

Query Match 38.8%; Score 71; DB 10; Length 174;
 Best Local Similarity 44.8%; Pred. No. 0.021;
 Matches 13; Conservative 7; Mismatches 2; Gaps 1;

OY 6 CRGQIQEONLRQCCQYIKQV--SGQGP 32
 DB 38 CRREFQAOHLRACQOMLHRQARQSGSGP 66

RESULT 14
 ID 042469 PRELIMINARY; PRT; 180 AA.
 AC 042469;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE NAPIN.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. TOPAS;
 RX MEDLINE-95161697; PubMed-7858212;
 RA Boullier K.A., Gines M.J., Demoor J.M., Huang B., Basczynski C.L.,
 Iyer V.N., Miki B.L.;
 RT "Expression of the BmNAP subfamily of napin genes coincides with the
 induction of Brassica microspore embryogenesis.";
 RL Plant Mol. Biol. 26:1711-1723(1994).
 DR EMBL: U04944; AAA81908.1;
 DR EMBL: U04943; AAA81907.1;
 DR InterPro: IPR003612; AAT.
 DR InterPro: IPR001768; Cereals-trypan-amylin.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; trypan-amylin.1.
 DR PRINTS: PR00496; NAPIN.
 DR ProDom: PD002498; Napin.1.
 DR SMART: SM00499; AAT; 1.
 SQ SEQUENCE 180 AA; 20342 MW; 327FC2A7D9838E3C CRC64;

Query Match 38.8%; Score 71; DB 10; Length 180;
 Best Local Similarity 44.8%; Pred. No. 0.022;
 Matches 13; Conservative 7; Mismatches 2; Gaps 1;

OY 6 CRGQIQEONLRQCCQYIKQV--SGQGP 32
 DB 47 CRREFQAOHLRACQOMLHRQARQSGSGP 75

RESULT 15
 ID 09S9F1 PRELIMINARY; PRT; 41 AA.
 AC 09S9F1;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE NAPIN SHORT CHAIN S4-CALMODULIN ANTAGONIST/CALCIUM-DEPENDENT PROTEIN
 KINASE SUBSTRATE.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE-96283790; PubMed-8679670;
 RA Neumann G.M., Condron R., Thomas I., Polya G.M.;
 RT "Purification and sequencing of multiple forms of Brassica napus seed
 napin small chains that are calmodulin antagonists and substrates for
 plant calcium-dependent protein kinase.";
 RL Biochim. Biophys. Acta 1295:23-33(1996).
 SQ SEQUENCE 41 AA; 4652 MW; 8BC6738503380553 CRC64;

Query Match 38.3%; Score 70; DB 10; Length 41;
 Best Local Similarity 41.9%; Pred. No. 0.0064;
 Matches 13; Conservative 7; Mismatches 7; Indels 4; Gaps 1;

Qy 6 CRGQIQEQONLRQCGEYIKQV----SGQGP 32
| | : | : | | | | : | | | |
Db 10 CRKEFQQAQHLRACCOQWTLHKQAMQFGSGSGP 40

Search completed: September 9, 2002, 12:45:37
Job time: 294 sec

4.3.1.1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:42:39 ; Search time 94.6 Seconds
(without alignments)
76.319 Million cell updates/sec

Title: US-09-913-351-5

Perfect score: 350
Sequence: 1 QERSLRGCCDHAKQMSQCR.....FRTANLPSCMGVSPEKRF 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: A-Geneseq.032802.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	350	100.0	65	21	AA15423
2	136	38.9	158	18	AAW23586
3	134	38.3	158	18	AAW23588
4	133.5	38.1	111	19	AAW6269
5	132.5	37.9	93	12	AA14111
6	132.5	37.9	110	15	AAW62762
7	132.5	37.9	111	19	AAW6271
8	132.5	37.9	111	19	AAW6264
9	132.5	37.9	140	10	AAW1891
10	131.5	37.6	171	15	AAW3579
11	130.5	37.3	111	19	AAW6267

12	130.5	37.3	111	19	AAW6268	Modified Brazil nu
13	130.5	37.3	111	19	AAW6270	Modified Brazil nu
14	130.5	37.3	149	19	AAW6272	Modified Brazil nu
15	130.5	37.3	149	19	AAW6273	Modified Brazil nu
16	129.5	37.0	111	19	AAW6266	Modified Brazil nu
17	129.5	37.0	111	19	AAW6265	Modified Brazil nu
18	121	34.6	68	21	AA15422	Castor bean 2S alb
19	117.5	33.6	149	19	AAW6274	Modified Brazil nu
20	112	32.0	164	10	AAW1892	Amino acid sequenc
21	112	32.0	164	10	AAW6144	Sequence encoded b
22	108	30.9	148	21	AAW58049	Puroindoline B pro
23	108	30.9	148	21	AAW58050	Serine substituted
24	106.5	30.4	155	18	AAW10569	Mablin II prepro
25	106.5	30.4	155	18	AAW23587	Mablin MBLI fro
26	104	29.7	166	21	AAW33600	Modified Ara h 2 a
27	104	29.7	166	22	AAW05035	Modified anaphylac
28	104	29.7	167	22	AAW04710	Modified anaphylac
29	101	28.9	156	20	AAW40973	Ara h 2 protein fr
30	101	28.9	157	18	AAW24164	Peanut allergen Ar
31	101	28.9	157	18	AAW24153	Peanut allergen Ar
32	101	28.9	157	20	AAW15245	Peanut allergen A
33	101	28.9	157	22	AAW04707	Anaphylactic anti
34	101	28.9	166	20	AAW40968	Recombinant Ara h
35	101	28.9	207	22	AAW82383	Peanut allergen Ar
36	99	28.3	175	14	AAW33390	Albumin-type prote
37	96.5	27.6	184	12	AAW41493	Napin 1-2 gene pro
38	96.5	27.6	184	15	AAW54981	Napin protein from
39	96.5	27.6	184	15	AAW54981	B. campestris nap
40	96.5	27.6	184	18	AAW15415	Napin encoded by c
41	96.5	27.6	184	19	AAW30526	Brassica campestri
42	96.5	27.6	184	22	AAW85931	B. campestris lamb
43	96.5	27.6	189	9	AAW81140	Sequence encoded b
44	95.5	27.3	184	19	AAW30670	Napin seed storage
45	94.5	27.0	189	15	AAW54980	Napin protein from

ALIGNMENTS

RESULT 1	AA15423	standard; peptide: 65 AA.
ID	AA15423	
AC	AA15423	
DT	17-JAN-2001	(first entry)
DE	Castor bean 2S albumin storage protein amino acids 194-258.	
KW	Antitumour; antimicrobial; immunostimulatory; glycoconjugate; mannose;	
KW	polysaccharide; mannose; galactose; castor bean; immune response; human;	
KW	2S albumin storage protein; animal; tumour necrosis factor; mononuclear;	
KW	phagocyte; granulocyte; macrophage; tumour; immunogen; Candida utilis.	
XX		
OS	Ricinus communis.	
XX		
FN	WO200050087-A1.	
XX		
PD	31-AUG-2000.	
XX		
PF	21-OCT-1999; 99WO-ES00338.	
XX		
PR	26-FEB-1999; 99ES-0000408.	
XX		
PA	(INFA-) IND FARM CANTABRIA SA.	
XX		
PI	Brieva Delgado A, Garcia Villarrubia V, Guerrero Gomez-Pamo A;	
PI	Piñel Ranieri JP, Gimenez Gallego G, Matji Tuduri JA;	
XX		
DR	WPI: 2000-558369/51.	
XX		
PT	New glycoconjugate, useful for treating immunological disorders,	
PT	comprises polysaccharide from Candida utilis and polypeptide from	

PT	Ricinus communis	-
XX	Claim 3; Page 30; 38pp; Spanish.	
CC	The invention relates to a novel glycoconjugate (A) comprising a	
CC	50-250 kDa polysaccharide (PS), containing one phosphate group per	
CC	5-25 monosaccharide units, and a polypeptide (PP) having a particular	
CC	consensus sequence. In PS, at least 40% of the monosaccharide residues	
CC	are mannose and the remainder are glucose and/or galactose. The main	
CC	chain comprises 1-6 links, with less than 60% 1,2-branching. The PP has	
CC	the consensus sequence:	
CC	ZacDbC(QERK)Z(LVIM)ZCC(Z'')(QEH)(LV)ZGCGZAZGZCZ(VILM)ZfCZg	
CC	where Z = any amino acid (aa); z' = hydrophobic aa; z'' = hydrophilic aa;	
CC	a = 3-8; b = 9-13; c = 15-39; d = 13-56; e = 15-26; f = 1-8; g = 1-11;	
CC	the parentheses indicate the preferred sequence. This sequence	
CC	corresponds to amino acids 194-258 of the Ricinus communis (castor bean)	
CC	2S albumin storage protein which is used to generate the glycoconjugate	
CC	of the invention, with a polysaccharide from Candida utilis. (A) modifies	
CC	the immune response in humans and animals, especially it reduces the	
CC	amount of tumour necrosis factor (TNF) produced and stimulates the	
CC	mononuclear phagocyte system and expands the granulocyte-macrophage	
CC	compartment. (A) are used for treating disorders of the immune system,	
CC	e.g. infections and tumours. (A) do not interfere with metabolic	
CC	processes in the liver and are active when given orally. PP significantly	
CC	increases the antigenic response to weakly immunogenic PS, including	
CC	induction of a T-dependent response.	
SQ	Sequence 65 AA:	
OY	Query Match 100.0%; Score 350; DB 21; Length 65;	
OY	Best Local Similarity 100.0%; Pred. No. 2.9e-37;	
OY	Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DB	1 GERSTGCCDHLHKOMOSCEGLRQALPQOQSOGLOJGDVEAFRTANLPSCMGVSP 60	
DB	1 gerslrgccdhkhkmgsgqrceglrqaiegqsggg19gdvfearttanlpmscvsp 60	
OY	61 TECRF 65	
OY		
DB	61 tegrff 65	
RESULT 2		
ID	AAM23586	
XX	AAM23586 standard; Protein: 158 AA.	
AC	AAM23586;	
DT	30-SEP-1997 (first entry)	
DE	Mablinln MBLI from Capparis masalkai.	
KW	MBL: sweetener; transgenic plant; recombinant protein; food;	
KW	beverage; animal feed; chewing gum; dental hygiene product;	
KW	pharmaceutical.	
OS	Capparis masalkai.	
FH	Key Location/Qualifiers	
FT	Peptide 1..39	
FT	/label= Signal	
FT	Protein 40..158	
FT	/label= MBLI	
FT	Region 40..70	
FT	/label= A_chain	
FT	Region 71..86	
FT	/note= "14 amino acids evidently lost when mature	
FT	dimer is formed"	
FT	Region 87..158	
FT	/label= B_chain	
NN	WO9700945-A1	

PD	09-JAN-1997.	
XX		
PF	21-JUN-1996;	96WO-US10669.
XX		
PR	23-JUN-1995;	95US-0000480.
XX		
PA	(UYHA-) UNIV HAMAIL.	
PI	Chen H, Hu Z, Sun SSM, Xiong L;	
DR	WPI: 1997-087372/08.	
XX	N-PSDB; AAT60775.	
PT	Mablinin nucleotide sequences - for prodn. of transgenic plants or	
PT	recombinant proteins, useful as sweeteners, partic. in foods	
XX		
PS	Disclosure; Fig 3; 38pp: English.	
XX		
CC	The present sequence represents the mablinin MBLI. The encoding cDNA can	
CC	modulate the expression of the gene encoding MBL in a cell or tissue.	
CC	CC Mablinin comes from the seeds of the Capparis masakal plant, and	
CC	is limited by its native availability, and so methods for producing	
CC	recombinant mablinin protein are being produced using chimeric genes	
CC	such as MBLII (see AAT60776). The products provide recombinant materials	
CC	for the production of the MBL protein in practical amounts, and for the	
CC	production of transgenic plants containing inherently sweet edible parts	
CC	by virtue of production of MBL in situ. The MBL protein can be used to	
CC	sweeten foods, beverages, animal feeds, chewing gum, dental hygiene	
XX	products and pharmaceuticals.	
SQ	Sequence 158 AA:	
Query Match	38.9%; Score 136; DB 18; Length 158;	
Best Local Similarity	43.5%; Pred. No. 1.8e-09;	
Matches 30; Conservative 13; Mismatches 22; Indels 4; Gaps 3.		
OY	1 QER--SLRCCDHLKQMOQSCRCBGLRQALIEQOQSOQIQG-QDYFEAFRTAANLPSMCG 57	
	: : : : : :	
DB	87 gqrgpalrlcncqlqgvnkpvcvplrlqraahqlyqsgleqprgyqlftraarlnpnlok 146	
OY	58 VSPF-ECRF 65	
	: :	
DB	147 ipavgrcqf 155	
RESULT 3		
AAW23588		
ID	AAW23588 standard; Protein; 158 AA.	
XX		
AC	AAW23588;	
XX		
TT	30-SEP-1997 (first entry)	
XX		
DE	Mablinin MBLIII from Capparis masakal.	
XX		
KW	MBL, sweetener; transgenic plant; recombinant protein; food;	
KW	beverage; animal feed; chewing gum; dental hygiene product;	
KW	pharmaceutical.	
XX		
OS	Capparis masakal.	
XX		
FT	Key	Location/Qualifiers
FT	Peptide	1..39
FT	Protein	/label= Signal
FT		40..158
FT	Region	/label= MBLIII
FT		40..70
FT	Region	/label= A_chain
FT		71..86
FT	/note= "14 amino acids evidently lost when mature	
FT	dimer is formed"	

```

FT Region 87..158
FT /label= B_chain
PN WO9700945-A1.
XX
XX 09-JAN-1997.
PD
XX 21-JUN-1996; 96WO-US10669.
PF
XX 23-JUN-1995; 95US-0000480.
PR
XX (UYHA-) UNIV HAWAII.
PA
XX Chen H, Hu Z, Sun SSM, Xiong L;
PI WPI; 1997-087372/08.
DR
XX Mabinlin nucleotide sequences - for prodn. of transgenic plants or
XX recombinant proteins, useful as sweeteners, partic. in foods
PT
XX Disclosure; Fig 3; 38pp; English.
PS
XX The present sequence represents the mabinlin MBL11. The encoding cDNA
XX can modulate the expression of the gene encoding MBL in a cell or
XX tissue. Mabinlin comes from the seeds of the Capparis masaikei plant,
XX and is limited by its native availability, and so methods for producing
XX recombinant mabinlin protein are being produced using chimeric genes
XX such as MBL11 (see AAT6076). The products provide recombinant materials
XX for the production of the MBL protein in practical amounts, and for the
XX production of transgenic plants containing inherently sweet edible parts
XX by virtue of production of MBL in situ. The MBL protein can be used to
XX sweeten foods, beverages, animal feeds, chewing gum, dental hygiene
XX products and pharmaceuticals.
SQ Sequence 158 AA:

Query Match 38.3%; Score 134; DB 18; Length 158;
Best Local Similarity 43.5%; Pred. No. 3.3e-09;
Matches 30; Conservative 12; Mismatches 23; Indels 4; Gaps 3;

QY 1 QER--SLRGCCHLKOMOSQCRCEGLRQAIEQQOSQGLQGDVFEAFRTAANLPSMCG 57
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 87 gqrgpalrlccnqlrqnvpvcvplrqaahqqlygqglegprqrvrrlfrtaarlnpnick 146
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 58 VSPF-ECRF 65
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 147 lpavgrcql 155

RESULT 4
AAW86269
ID AAW86269 standard; Protein; 111 AA.
XX
AC AAW86269;
XX
DT 01-MAR-1999 (first entry)
XX
DE Modified Brazil nut 2S albumin gene BN18 amino acid sequence.
XX
KW Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KW BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN19; AT2S1BN153W;
KW chimeric; dietary protein; essential amino acid.
XX
OS Bertholletia excelsa.
OS Synthetic.
XX
PN WO9845458-A1.
XX
PD 15-OCT-1998.
XX
PF 06-APR-1998; 98WO-US06673.
XX

```

```

PR 08-APR-1997; 97US-0042827.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA
XX Gutteridge S;
XX
XX WPI; 1998-609902/51.
DR N-PSDB; AAY80249.
XX
XX New modified Brazil Nut 2S albumin proteins - useful to provide
XX transgenic seed crops whose proteins have increased levels of
XX essential amino acids
PT
XX Claim 6; Fig 6; 80pp; English.
PS
XX The invention relates to modified Brazil Nut 2S albumin seed storage
XX proteins and nucleic acids encoding the modified proteins. The modified
XX Brazil nut 2S albumin proteins are rich in essential amino acids and the
XX encoding genes selected from a group comprising of BNCNSS, BN11, BN15,
XX BN17, BN18, BN19, BN153KW, AT2S1BN15, AT2S1BN19, AT2S1BN153W genes.
XX Chimeric genes comprising a nucleic acid fragment encoding a modified
XX Brazil nut 2S protein operably linked to suitable regulatory sequences,
XX integrated into the genome of a transformed plant can be used in
XX methods for increasing the essential amino acid content of seeds. The
XX products can be used to provide a human and animal dietary protein in
XX seed crop plants that is relatively high in essential amino acid content.
XX The present sequence represents the amino acid sequence of a modified
XX Brazil Nut 2S albumin gene BN18. The modified BN18 has all the 15 Arg
XX residues, Gly at position 105, and Ser at positions 44 and 107 of the
XX wild-type protein replaced with Lys residues
SQ Sequence 111 AA:

```

```

Query Match 38.1%; Score 133.5; DB 19; Length 111;
Best Local Similarity 35.4%; Pred. No. 2.5e-09;
Matches 23; Conservative 17; Mismatches 24; Indels 1; Gaps 1;

QY 2 ENSLRGCCHLKOMOSQCRCEGLR-QAIEQQOSQGLQGDVFEAFRTAANLPSMCGVSP 60
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 40 ephmkceceqlegmdesckceglkmmumkmgqemqpkqgmkkmlaenlpskcnlsp 99
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 61 TECRF 65
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 100 mkcpl 104

RESULT 5
AAR14111
ID AAR14111 standard; Protein; 93 AA.
XX
AC AAR14111;
XX
DT 10-DEC-1991 (first entry)
XX
DE Brazil nut 12 kD sulfur-rich seed storage protein, BZ-3.
XX
KW Expression cassette; tryptophan; foodstuff; bovine somatotropin;
KW growth hormone.
XX
OS Synthetic.
XX
OS Key Location/Qualifiers
FH Region 24..89
FT /label= BZ-2
FT Region 28..75
FT /label= BZ-1
XX
XX WO9113993-A.
XX
XX 19-SEP-1991.
XX
XX 14-FEB-1991; 91WO-US00887.
XX

```

05-MAR-1990; 90US-0489175.
(UPJO) UPJOHN CO.
Slightom JL, Chae PP;
MPI: 1991-295647/40.
N-PSDB; AAQ13801.
Seed specific expression cassettes - having specific gene
regulatory elements for prodn. of proteins in plant seed storage
bodies
Disclosure; Page 38, 58pp; English.
Three oligomers were synthesised representing portions of the Brazil
nut coding DNA. The first oligomer (tag b) encodes the major portion
of the 9 kd subunit (B2-1), which contains 35% sulfur amino acids for
the 48 amino acid peptide. The second oligomer (tag a) consisting of
the complete 9 kd subunit (B2-2) encodes 70 amino acids contg. 28.5%
sulfur amino acids. The third oligomer is the complete 12 kd subunit
(B2-3) which encodes 93 amino acids contg. 28% sulfur amino acids.
The fragments were complemented, purified, and had NcoI sites added on
both 5' and 3' ends using PCR amplification. Each inframe NcoI site
yields the codon TGC which encodes tryptophan, making resulting plant
seed a more nutritionally complete foodstuff. The fragments were
subjected to digestion with NcoI and cloned into NcoI cut expression
cassettes. Plasmids contg. such expression cassettes were transferred
into vectors used for Agrobacterium systems or used with microprojectile
bombardment of plant tissues.
The seed specific expression cassette is used for the prodn. of
bodies.
See also AAQ13798.
Sequence 93 AA:

Query Match	37.9%;	Score 132.5;	DB 12;	Length 93;
Best Local Similarity	39.7%;	Pred. No. 2.7e-09;		
Matches 25;	Conservative 10;	Mismatches 27;	Indels 1;	Gaps 1.
QY	2	ERSLNGCCDHLKOMOSQCRCEGLR-QAIEQOOSQGLQSDVFEAFRTANLPMSGVSP	60	
		: :	: : :	:
Db	29	ephmseeceqlegmdeserccglrmmmmmqgemqpgyqgmrrmlaenlpsrcnlsp	88	
QY	61	TEC	63	
Db	89	mrc	91	
RESULT	6			
AA62762				
ID	AA62762	standard; Protein; 110 AA.		
XX				
AC	AA62762;			
XX				
DT	07-JUL-1995	(first entry)		
XX				
DE	Synthetic 12 Kd precursor protein of Brazil nut.			
KX	Stabilisation; toxic; peroxisome; yeast; food additive; brazil nut.			
XX	Synthetic.			
OS				
XX	WO9424289-A.			
PN				
XX				
PD	27-OCT-1994;			
XX				
PF	19-APR-1994;	94WO-FR00438.		
XX				
PR	19-APR-1993;	93FR-0004583.		

XX	(EURO-) EUROLYSINE.
PA	
XX	
PI	Ito H, Labat N, Nicaud J, Pardo D, Raynal A, Sugimoto S;
XX	
DR	WPI: 1994-341868/42.
DR	N-PSDB; AAQ73069.
XX	
PT	Improving stability of protein or peptide in yeast - by
PT	expressing with a peroxisomal targeting sequence, allowing
PT	accumulation of unstable or toxic products, and new yeast strains
XX	useful as food additives
XX	
PS	Claim 8; Figure 1; 40pp; French.
XX	
CC	Stabilisation of a non-peroxisomal protein or peptide that is
CC	unstable in, or toxic for, a yeast and made by expressing a
CC	heterologous DNA sequence in the yeast, comprises expressing it with
CC	a sequence that targets to peroxisomes to stabilise/detoxify in the
CC	yeast peroxisomes. The method is especially used when the non-
CC	peroxisomal protein is the 2S storage protein (this sequence) of
CC	Brazil nuts or the synthetic peptides E (see AAR62763) and E4 (see
CC	AAR62764). Yeast expressing these proteins are useful as food
CC	additives (having a high content of essential amino acids).
XX	
Sequence	110 AA;
50	

	Query Match	37.9%	Score 132.5;	DB 15;	Length 110;
Dc	Best Local Similarity	39.7%;	Pred. No.	3.3e-09;	
	Matches 25; Conservative 10; Mismatches 27; Indels 1; Gaps				
Qy	2 ERSRGCCDHLKQMOSOCREGLR-QAIEEQSQSGQLGDVFEEAFRTAAALPSMGVSP 60 : : : : : : : : : : : : : Db ephmsceeqeglegndescrceglrlmmumnmqgemgprgegmrrmmlaeiniprcalpsl 100				
Dd	61 TEC 63 : : : : : : : : : : : : : Db 101 mrc 103				
	RESULT 7				
ID	AAW86271 standard; Protein; 111 AA. AAW86271: 01-MAR-1999 (first entry)				
Xx	Modified Brazil nut 2S albumin gene BN153KW amino acid sequence.				
KM	Brazil Nut 2S albumin; seed storage protein; modified; BNCSS; BN11; KM BN15; BN17; BN18; BN15; ATZSIBN15; ATZSIBN19; ATZSIBN13W; XM chimeric; dietary protein; essential amino acid.				
OS	Bertholletia excelsa.				
FH	Synthetic.				
Key	Location/Qualifiers				
FT	/label= EAK				
FT	/note= "wild-type Glu is replaced with Lys"				
FT	/label= R6K				
FT	/note= "wild-type Arg is replaced with Lys"				
FT	/label= R11K				
FT	/note= "wild-type Arg is replaced with Lys"				
FT	/label= S16K				
FT	/note= "wild-type Ser is replaced with Lys"				
FT	/label= R19K				

FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 23	/label= R23K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 36	/label= R36K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 37	/label= R37K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 41	/label= P41K
FT		/note= "wild-type Pro is replaced with Lys"
FT	Misc-difference 42	/label= W42H
FT		/note= "wild-type Trp is replaced with His"
FT	Misc-difference 58	/label= R58K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 63	/label= R63K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 68	/label= R68K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 77	/label= R77K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 82	/label= R82K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 83	/label= R83K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 86	/label= R86K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 94	/label= R94K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference 101	/label= R101K
FT		/note= "wild-type Arg is replaced with Lys"
XX	W09845458-A1.	
XX	15-OCT-1998.	
XX	06-APR-1998;	98WO-US06673.
XX	08-APR-1997;	97US-0042827.
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.	
XX	Gutteridge S;	
PI	WPI: 1998-609902/51.	
XX	N-PSDB: AAV80251.	
DR		
XX		
XX	New modified Brazil Nut 2S albumin proteins - useful to provide	
PT	transgenic seed crops whose proteins have increased levels of	
PT	essential amino acids	
XX		
PS	Claim 6; Fig 8; 80pp; English.	
XX		
CC	The invention relates to modified Brazil Nut 2S albumin seed storage	
CC	proteins and nucleic acids encoding the modified proteins. The modified	
CC	Brazil nut 2S albumin proteins are rich in essential amino acids and the	
CC	encoding genes selected from a group comprising of BNCNS, BN11, BN15,	
CC	BN17, BN18, BN19, BN15KW, AT251BN15, AT251BN19, AT251BN15W genes.	
CC	Chimeric genes comprising a nucleic acid fragment encoding a modified	
CC	Brazil nut 2S protein operably linked to suitable regulatory sequences,	
CC	integrated into the genome of a transformed plant can be used in	

CC methods for increasing the essential amino acid content of seeds. The
CC products can be used to provide a human and animal dietary protein in
CC seed crop plants that is relatively high in essential amino acid content.
CC The present sequence represents the amino acid sequence of a modified
CC Brazil Nut 2S albumin gene BN153KW.

XX Sequence 111 AA:

SQ

Query Match 37.9% Score 132.5 : DB 19 ; Length 111;
Best Local Similarity 35.4% ; Pred. No. 3.4e-09;
Matches 23; Conservative 17; Mismatches 24; Indels 1; Gaps 1;

OY 2 ERSLRGCCDHLKMSOCRCGLR-QAIEOQSGOLGODVFEAFRTAANLPSCGVSP 60
 | : | 11 : | : | 1 : |||| : :: | | : | : | : | 1 : | |
DB 40 ekvmecceccqlegmdesckcgLkmmnmkmgemgpkxgegmkmklaelnpskcnlsp 99

OY 61 TECRF 65
 : | |
DB 100 mkcpf 104

RESULT 8
AAW86264
ID AAW86264 standard; Protein; 111 AA.

XX
AC AAW86264;
XX
DT 01-MAR-1999 (first entry)

DE Wild-type Brazil nut 2S albumin seed storage protein.

XX
XX
KM Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KW BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN153W;
KK chimeric; dietary protein; essential amino acid.

XX Bertholletia excelsa.
OS
PN WO9845458-A1.
XX
PD 15-OCT-1998.
XX
PF 06-APR-1998; 98WO-US06673.
XX
PR 08-APR-1997; 97US-0042827.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PI
PI Guttridge S;
XX
DR WPI: 1998-609902/51.
DR N-PSDB; AAV80240.
XX

Pt New modified Brazil Nut 2S albumin proteins - useful to provide
PT transgenic seed crops whose proteins have increased levels of
PT essential amino acids

XX Example 1; Fig 1; 80pp; English.

PS
XX This represents a wild-type Brazil Nut 2S albumin seed storage
CC protein. The invention relates to modified Brazil Nut 2S albumin
CC proteins that are rich in essential amino acids and nucleic acids
CC encoding the modified proteins. The modified Brazil nut 2S albumin are
CC selected from a group comprising of BNCNSS, BN11, BN15, BN17, BN18, BN19,
CC BN153KW, AT2S1BN15, AT2S1BN19, AT2S1BN153W genes. Chimeric genes
CC comprising a nucleic acid fragment encoding a modified Brazil nut 2S
CC protein operably linked to suitable regulatory sequences, integrated into
CC the genome of a transformed plant can be used in methods for
CC increasing the essential amino acid content of seeds. The products can be
CC used to provide a human and animal dietary protein in seed crop plants
CC that is relatively high in essential amino acid content.

XX Sequence 111 AA:
SQ

```

Query Match Similarity 37.9%; Score 132.5; DB 19; Length 111;
Best Local Similarity 39.7%; Pred. No. 3.4e-09;
Matches 25; Conservative 10; Mismatches 27; Indels 1; Gaps 1.

QY      2 ERSLRCCDHLKQMSQCRCEGLR-QALIEQSQSQGLQGDVFEAFRTANLPSMGVSP 60
      1 : 11 : 1 : 111111 : 11 : 1 : 1 : 11 : 11
Db      40 ephmesceqlegmdscrcceglrmnmrmqgemqprgqgmrrmrmlaenlpsrcnla 99
QY      61 TEC 63
      100 mrc 102

RESULT 9
AAP91891
ID AAP91891 standard; protein; 140 AA.
XX
XX AAP91891;
XX
XX 29-APR-1990 (first entry)
XX
XX
XX Part of the sequence of the Brazil nut 2S-albumin as encoded in the
DE PBN2S1 plasmid.
XX
XX 2S-albumin; brazil nut; PBN2S1; storage protein gene;
KM heterologous polypeptide.
XX
XX Brazil nut.
OS
XX
XX
XX Key Location/Qualifiers
FH Peptide 1..30
FT /note="signal peptide"
FT protein 31..38
FT /note="mature small subunit"
FT region 39..43
FT /note="processing site"
FT protein 44..136
FT /note="mature large subunit"
FT
XX
XX W08903887-A.
XX
XX
XX 05-MAY-1989.
XX
XX 20-OCT-1988; 88WO-EP00944.
XX
XX 20-OCT-1987; 87EP-0402348.
XX
XX (PLAN ) PLANT GENETIC SYST.
XX
XX Vandekerckhove JS, Krebbers E, Botterman J, Leemans J;
XX
XX WPI: 1989-150783/20.
XX
XX N-PSDB: AAN91699.
XX
XX Recombinant DNA expression in plants
PT - using modified storage protein genes for expressing
PT heterologous polypeptide(s) in the seeds
XX
XX Fig 4; ; 121pp; English.
XX
XX The entire 2S-albumin storage protein precursor including
XX signal peptide. It is to be inserted into plants under the control of
XX a seed-specific promoter and expressed at high levels only or mostly
XX in the seed forming stage and produced mostly in the seeds.
XX
XX Sequence 140 AA;

```

Oy		2	EESLFGCCDHHKQMOOSRCGELR-QAIEQQOQSOGQGDVFEAFPTANLPSCGVSP	60
Dd		69	ephmsecceqlegmdescrceqlrrmmmrngqeampgreqmrrrmrlaelnlpshrnlsp	128
Oy		61	TEC 63	
		↓		
Db		129	mrc 131	
		RESULT_10		
		AAR53579		
ID		AAR53579	standard; Protein; 171 AA.	
XX				
AC		AAR53579;		
XX				
DT		25-OCT-1994	(first entry)	
XX				
DE		Synthetic 2S seed storage protein.		
XX				
KW		Nutrition; non-conserved DNA; assembly; crude; degenerate; oligonucleotides.		
XX				
OS		Synthetic.		
XX				
FH		Key	Location/Qualifiers	
FT		Protein	39..74	
FT			/note= "small subunit"	
FT		Protein	85..170	
FT			/note= "large subunit"	
FT		Peptide	1..38	
FT			/note= "signal peptide"	
FT		Peptide	75..84	
FT			/note= "linker excised as protein folds"	
FT		Misc-difference	171	
FT			/note= "excised at protein maturity"	
XX				
PN		WO9410315-A.		
XX				
PD		11-MAY-1994.		
XX				
PF		22-OCT-1993;	93MO-USI0090.	
XX				
PR		23-OCT-1992;	92US-095664.	
XX				
PA		(PION-) PIONEER HT-BRED INT INC.		
XX				
PI		Ballo B;		
XX				
DR		WPI; 1994-167470/20.		
XX				
DNR		N-PSSB; AAO63710.		
XX				
PT		Improving the content of an amino acid in a seed storage protein to enhance nutritional value - by replacing non-conserved DNA with DNA encoding the amino acid		
XX				
PS		Disclosure; Fig 1; 33pp; English.		
XX				
CC		The sequence is that of a 2S seed storage protein with increased lysine content. The protein is synthesised from a set of partial cDNA sequences capable of being assembled in a complementary over- lapping relationship to provide the complete DNA. The protein is translated as the straight chain shown and undergoes post translational modification to give the mature protein which has two subunits joined by a di-sulphide bond.		
XX				
Sequence		171 AA;		
10				

Query Match	37.98;	Score 132.5;	DB 10;	Length 140;
Best Local Similarity	39.78;	Pred. No. 44e-09;		
Matches	25;	Conservative	10;	Mismatches 27;
				Indels 1;
				Gaps 1

Query Match	37.6%;	Score 131.5;	DB 15;	Length 171;
Best Local Similarity	44.1%;	Pred. No. 7.6e-09;		
Matches	30;	Conservative	14;	Mismatches 17;
			Indels	7;
			Gaps	4;


```

RESULT 13
AAW86270
ID AAW86270 standard; Protein; 111 AA.
XX
AC AAW86270;
XX
DT 01-MAR-1999 (first entry)
XX
DE Modified Brazil nut 2S albumin gene BN19 amino acid sequence.
XX
XX Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KM BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN153W;
KW chimeric; dietary protein; essential amino acid.
XX
OS Bertholletia excelsa.
XX
XX Synthetic.
XX
FH Key
FT Misc-difference 4 Location/Qualifiers
FT /label= E28K
FT /note= "wild-type Glu is replaced with Lys"
FT Misc-difference 11
FT /label= R11K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 16
FT /label= R16K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 19
FT /label= R19K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 23
FT /label= R23K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 27
FT /label= E27M
FT /note= "wild-type Glu is replaced with Met"
FT Misc-difference 28
FT /label= E28K
FT /note= "wild-type Glu is replaced with Lys"
FT Misc-difference 36
FT /label= R36K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 37
FT /label= R37K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 58
FT /label= R58K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 63
FT /label= R63K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 68
FT /label= R68K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 77
FT /label= R77K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 82
FT /label= R82K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 83
FT /label= R83K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 86
FT /label= R86K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 94
FT /label= R94K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 101
FT /label= R101K

```

```

FT /note= "wild-type Arg is replaced with Lys"
XX
XX WO9845458-A1.
XX
XX 15-OCT-1998.
XX
XX 06-APR-1998; 98WO-US06673.
XX
XX 08-APR-1997; 97US-0042827.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E. I.
XX
XX Gutteridge S;
XX
XX WPI; 1998-609902/51.
XX
XX N-PSDB; AAW80250.
XX
XX New modified Brazil Nut 2S albumin proteins - useful to provide
PT transgenic seed crops whose proteins have increased levels of
PT essential amino acids
XX
XX Claim 6; Fig 7; 80pp; English.
XX
XX The invention relates to modified Brazil Nut 2S albumin seed storage
CC proteins and nucleic acids encoding the modified proteins. The modified
CC Brazil nut 2S albumin proteins are rich in essential amino acids and the
CC encoding genes selected from a group comprising of BNCNSS, BN11, BN15,
CC BN17, BN18, BN19, BN153KW, AT2S1BN15, AT2S1BN153W genes.
CC Chimeric genes comprising a nucleic acid fragment encoding a modified
CC Brazil nut 2S protein operably linked to suitable regulatory sequences,
CC integrated into the genome of a transformed plant can be used in
CC methods for increasing the essential amino acid content of seeds. The
CC products can be used to provide a human and animal dietary protein in
CC seed crop plants that is relatively high in essential amino acid content.
CC The present sequence represents the amino acid sequence of a modified
CC Brazil Nut 2S albumin gene BN19.
XX
XX Sequence 111 AA:
SQ
Query Match 37.3%; Score 130.5; DB 19; Length 111;
Best Local Similarity 35.4%; Pred. No. 6e-09;
Matches 23; Conservative 16; Mismatches 25; Indels 1; Gaps 1;
QY 2 ERSLRGCCDHKKQMSQRCCEGLR-QAIEQQSGGQLOGDVFPAFRANLPSCGVSP 60
DB 40 ephmseceqlegndesckceglkmmnmknqgeemqpgkmmkkaenlpsknlsp 99
QY 61 TECRF 65
DB 100 mkcpf 104

```

```

RESULT 14
AAW86272
ID AAW86272 standard; Protein; 149 AA.
XX
AC AAW86272;
XX
DT 01-MAR-1999 (first entry)
XX
DE Modified Brazil nut 2S albumin gene AT2S1BN15 amino acid sequence.
XX
XX Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KM BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN153W;
KW chimeric; dietary protein; essential amino acid.
XX
XX Bertholletia excelsa.
XX
XX Synthetic.
XX
XX WO9845458-A1.
XX
XX 15-OCT-1998.

```


• • • • •

Dd	87	QQRGALRLCCNQJLRQVYKKPCVCEVLNQAAHQJLYQGIESPQVRYQLFRARNLPIK	146
Qy	58	VSPT-ECRF	65
	:	1:1	
Dd	147	IPAVGRQGF	155

```

RESULT 2
US-08-670-186-6
Sequence 6, Application US/08670186
Patent No. 5859343
GENERAL INFORMATION:
APPLICANT: SUN, SAMUEL S.M.
APPLICANT: XIONG, LIWEN
APPLICANT: HU, ZHONG
APPLICANT: CHEN, HANG
TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,186
FILING DATE: 21-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 23461-20007.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 867-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSH
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
Type: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-670-186-6

```

Query Match	38.3%	Score 134	DB 2:	Length 158;
Best Local Similarity	43.5%	Pred. No. 1.3e-09;		
Matches 30;	Conservative 12;	Mismatches 23;	Indels 4;	Gaps 3

QY 1 QER--SLRGCDHLKQMOSQCCEGLRRAIEQQDSOGLOG-QDYEEAFRTAANLPSMCG 57
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 87 QQRRPALRLCCNQLRVYNKPCVCPLRLAAHQQLYGGQIEGRPRVYRLLFRARNLPNCK 146

QY	58	VSPT-ECRF	65
		:	1:1
Db	147	IPAVGRGQF	155

RESULT 3
US-08-670-186-2
Sequence 2, Application US/08670186
Patent No. 5853343
GENERAL INFORMATION:
APPLICANT: SUN, SAMUEL S.M.
APPLICANT: XIONG, LIMEN
APPLICANT: HU, ZHONG
APPLICANT: CHEN, HANG

```

1  TITLE OF INVENTION:  RECOMBINANT SWEET PROTEIN BABINLIN
2  NUMBER OF SEQUENCES:  10
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE:  MORRISON & FOERSTER
5  STREET:  2000 PENNSYLVANIA AVE NW, STE. 5500
6  CITY:  WASHINGTON
7  STATE:  DC
8  COUNTRY:  USA
9  ZIP:  20006-1888
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE:  floppy disk
12 COMPUTER:  IBM PC compatible
13 OPERATING SYSTEM:  PC-DOS/MS-DOS
14 SOFTWARE:  Patent in Release #1.0, Version #1.30
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER:  US/08/670,186
17 FILING DATE:  21-JUN-1996
18 CLASSIFICATION:  435
19 ATTORNEY/AGENT INFORMATION:
20 NAME:  MURASHIGE, KATE H.
21 REGISTRATION NUMBER:  29,959
22 REFERENCE/DOCKET NUMBER:  23461-20007.00
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE:  (202) 887-1500
25 TELEFAX:  (202) 822-0168
26 TELE:  90-4030 MRSFORSMW
27 INFORMATION FOR SEQ ID NO:  2:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH:  155 amino acids
30 TYPE:  amino acid
31 TOPOLOGY:  linear
32 MOLECULE TYPE:  protein
33 US-08-670-186-2

```

Query Match	30.4%	Score 106.5;	DB 2;	Length 155;
Best Local Similarity	39.3%	Pred. No. 3.7e-06;		
Matches 22;	Conservative 11;	Mismatches 22;		Indels 1; Gaps 1;

```
QY      4 SINGCCDHLKQMOSQCREGLRQAIEQQOOSGGLOG-QDVFEAFRTAANLPSMCGV 58
       :|||:|||:|||||:||||:|||||:|||||:|||||:
Db     88 ALRQCCNQLRQVDPRVCVPVLRLQAAGQVLQRQILQGFPQRLRLIEDARNLPNICNI 143
```

RESULT 4
US-08-453-924-3
Sequence 3, Application US/08453924
Patent No. 5608152
GENERAL INFORMATION:
APPLICANT: Kridl, Jean C.
APPLICANT: Knauf, Vic C.
TITLE OF INVENTION: Seed-Specific Transcriptional Regulation
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weill, Gotshal & Manges
STREET: 2882 Sand Hill Road, Ste. 280
CITY: Menlo Park
STATE: CA
COUNTRY: U.S.A.
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453, 924
FILING DATE: 30-MAY-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/742, 834
FILING DATE: 08-AUGUST-1991
PRIORITY APPLICATION DATA:

Query Match	22.6%;	Score 79;	DB 1;	Length 123;
Best Local Similarity	32.3%;	Pred. NO. 0.0081;		
Matches	21;	Conservative	10;	Mismatches 26;
			Indels	8;
			Gaps	3;

QY	QY	QY	QY
2	36	54	96
ERSLRGGCCHLKQMOSO	EAVLRDCCQGLADINNEM	SMCGV	EVCKV
-ORCEGLRQATEO	-RCGDDLSMRLRAYQEL	58	100
-QOSOGLOGDVFEAR	-GELGVEGNEVLPGCR		
-----TANLP	-----KVMKLTAAVP		
53	95		

```

RESULT 7
US-08-782-177-5
Sequence 5; Application US/08782177
Patent No. 5789380
GENERAL INFORMATION:
APPLICANT: MIYAZAKI, TOSHIYUKI
APPLICANT: MORIMOTO, TOSHIOHISA
APPLICANT: MURAYAMA, RYUJI
APPLICANT: TAKASE, SACHIKO
APPLICANT: GODA, TOSHINO
TITLE OF INVENTION: AGENTS FOR INHIBITING ACCUMULATION OF
TITLE OF INVENTION: VISCERAL FAT
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/782,177
FILING DATE: 14-JAN-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 23445/1996
FILING DATE: 18-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1327-055-0X
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-782-177-5

```

Query Match	22.6%	Score	79	DB	1	Length	123
Best Local Similarity	32.3%	Pred. No.	0.0081				
Matches	21	Conservative	10	Mismatches	26	Indels	8
						Gaps	3

0y 54 SMCV 58

Db 96 EVCKV 100

```

1      RESULT      8
2      US-08-618-911-6
3      Sequence 6, Application US/08618911
4      Patent No. 5850016
5      GENERAL INFORMATION:
6      APPLICANT: Jung, Rudolf
7      APPLICANT: Hastings, Craig
8      APPLICANT: Coughlan, Sean
9      APPLICANT: Hu, David
10     TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN
11     TITLE OF INVENTION: SEEDS
12     NUMBER OF SEQUENCES: 13
13     CORRESPONDENCE ADDRESS:
14     ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
15     STREET: 700 Capital Square, 400 Locust Street
16     CITY: Des Moines
17     STATE: Iowa
18     COUNTRY: USA
19     ZIP: 50309
20     COMPUTER READABLE FORM:
21     MEDIUM TYPE: Floppy disk
22     COMPUTER: IBM PC compatible
23     OPERATING SYSTEM: PC-DOS/MS-DOS
24     SOFTWARE: Patent Release #1.0, Version #1.30
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER: US/08/618,911
27     FILING DATE: Concurrently herewith
28     ATTORNEY/AGENT INFORMATION:
29     NAME: Simon, Soma
30     REGISTRATION NUMBER: 37,444
31     REFERENCE/DOCKET NUMBER: 365-US
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: (515) 248-4800
34     TELEFAX: (515) 248-4844
35     INFORMATION FOR SEQ ID NO: 6:
36     SEQUENCE CHARACTERISTICS:
37     LENGTH: 158 amino acids
38     TYPE: amino acid
39     TOPOLOGY: linear
40     MOLECULE TYPE: protein
41     US-08-618-911-6

```

Query Match	21.9%	Score	76.5	DB	2	Length	158
Best Local Similarity	24.6%	Pred. No.	0.023				
Matches	15	Conservative	20	Mismatches	23	Indels	3
						Gaps	2

Dy 1 QESLRGCCDHLKQMS-QCREGLRQLEQQQSOGLOGDVFEAFRTAANIPLSMGYS 59
 :: :: : :: :: : :: : :: : :: :
Db 90 EEGQMCKCTEMSELSPKCCKALQKTINQSE--ELLEKENKKMEKLMLNLTATMCRFG 147

QY	60 P	60
Db	148 P	148

```

RESULT      9
US-08-618-911-4
? Sequence 4, Application US/08618911
? Patent No. 5850016
?
? GENERAL INFORMATION:
?
? APPLICANT: Jung, Rudolf
?
? APPLICANT: Hastings, Craig
?
? APPLICANT: Coughlan, Sean
?
? APPLICANT: Hu, David
?
? TITLE OF INVENTION:  ALTERATION OF AMINO ACID COMPOSITIONS IN
?
? TITLE OF INVENTION:  SEEDS
?
? NUMBER OF SEQUENCES: 13
?
? CORRESPONDENCE ADDRESS:

```



```

; Patent No. 5789380
; GENERAL INFORMATION:
; APPLICANT: MIYAZAKI, TOSHIYUKI
; APPLICANT: MORIMOTO, TOSHIHISA
; APPLICANT: MURAYAMA, RYUJI
; APPLICANT: TAKASE, SACHIKO
; APPLICANT: GODA, TOSHINAO
; TITLE OF INVENTION: AGENTS FOR INHIBITING ACCUMULATION OF
; TITLE OF INVENTION: VISCERAL FAT
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,177
; FILING DATE: 14-JAN-1997
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 23445/1996
; FILING DATE: 18-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1327-055-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-782-177-1

```

```

Query Match 20.9%; Score 73; DB 1; Length 124;
Best Local Similarity 27.7%; Pred. No. 0.047;
Matches 18; Conservative 10; Mismatches 27; Indels 10; Gaps 2;

```

```

QY 2 ERLRCCDHLKOMOSQCRCEGLRAIEQOOSGGLQ---GQDVFEAF-----TAAN 51
| | | | | : | | | | | : | | | | | : | | | | |
DB 35 EAVLRDCCQDLADISWRCGALYSMLDSMYKHEGVQEGAGTGAPPCRRREVVKLTAA 94
| | | | | : | | | | | : | | | | | : | | | | |
QY 52 LPSC 56
| | | | | : | | | | | : | | | | | : | | | | |
DB 95 ITAVC 99
| | | | | : | | | | | : | | | | | : | | | | |

```

```

RESULT 15
US-08-782-177-3
; Sequence 3, Application US/08782177
; Patent No. 5789380
; GENERAL INFORMATION:
; APPLICANT: MIYAZAKI, TOSHIYUKI
; APPLICANT: MORIMOTO, TOSHIHISA
; APPLICANT: MURAYAMA, RYUJI
; APPLICANT: TAKASE, SACHIKO
; APPLICANT: GODA, TOSHINAO
; TITLE OF INVENTION: AGENTS FOR INHIBITING ACCUMULATION OF
; TITLE OF INVENTION: VISCERAL FAT
; NUMBER OF SEQUENCES: 5

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,177
; FILING DATE: 14-JAN-1997
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 23445/1996
; FILING DATE: 18-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1327-055-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ. ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-782-177-3

```

```

Query Match 20.9%; Score 73; DB 1; Length 124;
Best Local Similarity 29.2%; Pred. No. 0.047;
Matches 19; Conservative 9; Mismatches 27; Indels 10; Gaps 2;

```

```

QY 2 ERLRCCDHLKOMOSQCRCEGLRAIEQO-OSGGLQGGDVFEAF-----TAAN 51
| | | | | : | | | | | : | | | | | : | | | | |
DB 35 EAVLRDCCQDLADISWRCGALYSMLDSMYKHEGAGTGAPPCRRREVVKLTAA 94
| | | | | : | | | | | : | | | | | : | | | | |
QY 52 LPSC 56
| | | | | : | | | | | : | | | | | : | | | | |
DB 95 ITAVC 99
| | | | | : | | | | | : | | | | | : | | | | |

```

```

Search completed: September 9, 2002, 12:44:14
Job time: 236 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:43:30 ; Search time 45.25 Seconds
(without alignments)
138.029 Million cell updates/sec

Title: US-09-913-351-5

Perfect score: 350

Sequence: 1 QERSLRGCDHLKQMOSQCR.....FRYANLPSCGVSPTGCRF 65

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	350	100.0	258	1	RZCS
2	155.5	44.4	70	2	seed storage pr
3	147	42.0	154	2	seed storage prote
4	135	38.6	104	2	2S albumin - Brazil
5	133	38.0	100	2	2S albumin I-1 - Yun
6	133	38.0	104	2	2S albumin IV - Yun
7	132.5	37.9	146	2	2S albumin III - Yun
8	130	37.1	141	2	2S seed storage pr
9	119.5	34.1	95	2	2S albumin precurs
10	113.5	32.4	165	2	2S seed storage pr
11	112.5	32.1	139	2	2S seed storage prote
12	112.5	32.1	139	2	albumin 2S storage
13	112	32.0	164	1	albumin 2S storage
14	110.5	31.6	323	2	2S albumin I precu
15	108	30.9	148	2	2S albumin - commo
16	106.5	30.4	72	2	purindoline-b pre
17	106.5	30.4	155	2	2S albumin II chain
18	104.5	29.9	178	1	2S albumin II precu
19	104.5	29.9	178	2	2S albumin II precu
20	104.5	29.9	178	2	2S albumin II precu
21	104.5	29.9	178	2	2S albumin II precu
22	104	29.7	162	2	2S storage protein
23	104	29.7	162	2	2S storage protein
24	102.5	29.3	172	2	2S seed storage pr
25	102	29.1	124	2	2S-like storage pr
26	101.5	29.0	178	2	2S-like storage pr
27	101.5	29.0	178	2	2S-like storage pr
28	100.5	28.7	92	2	2S storage protein
29	99.5	28.4	162	2	2S storage protein

30	99	28.3	162	2	S49259	albumin 4 - easter
31	98.5	28.1	186	2	A23802	napin precursor (9
32	98	28.0	152	2	PS0427	napin AH1 precursor
33	97	27.7	174	2	PS0425	napin BA3 precursor
34	96.5	27.6	180	2	S10018	napin (clone BgNA
35	96.5	27.6	180	2	S52025	napin (clones BNM
36	96.5	27.6	327	2	JS0402	gamma-gliadin prec
37	95.5	27.3	155	2	S51770	grain softness pro
38	95.5	27.3	164	2	S48186	grain softness pro
39	95	27.1	145	2	PC1247	Sin a I allergen 2
40	95	27.1	166	1	NMMU4	2S albumin 4 precu
41	94.5	27.0	164	2	S48187	grain softness pro
42	94.5	27.0	170	1	NMMU2	2S albumin 2 precu
43	92	26.3	80	2	B23617	conglutinin delta-2
44	91.5	26.1	186	1	WMR219	19k globulin precu
45	91.5	26.1	186	2	JC4784	alpha-globulin pre

ALIGNMENTS

RESULT 1

RZCS

2S seed storage protein precursor - castor bean

N:Alternate names: 2S albumin precursor

C:Species: Ricinus communis (castor bean)

C:Date: 14-Nov-1983 #sequence; revision 08-Feb-1996 #text; change 18-Jun-1999

C:Accession: S11499; S11500; S11501; S11502; S27221; A01328; S27222

R:Irwin, S.D.; Lord, J.M.

Nucleic Acids Res. 18, 5890, 1990

A:Title: Nucleotide sequence of a Ricinus communis 2S albumin precursor gene.

A:Reference number: S11499; MUID:91016940

A:Accession: S11499

A:Molecule type: DNA

A:Residues: 1-258 <IRW>

A:Cross-references: EMBL:X54158; NID:921067; PIDN:CA38097.1; PID:921068

A:Note: the authors translated the codon CTC for residue 14 as Phe, CCA for residue 7

R:Irwin, S.D.; Keen, J.N.; Findlay, J.B.C.; Lord, J.M.

Mol. Gen. Genet. 222, 400-408, 1990

A:Title: The Ricinus communis 2S albumin precursor: a single preproprotein may be pro

A:Reference number: S11500; MUID:91109729

A:Accession: S11500

A:Molecule type: mRNA

A:Residues: 1-13, 'F', '15-73, 'T', '75-258 <IR2>

A:Experimental source: clone 1494

A:Accession: S11501

A:Molecule type: mRNA

A:Residues: 'M', '4', 'LS', '7-13, 'F', '15-21 <IRF>

A:Experimental source: clone 10a12

A:Accession: S27221

A:Molecule type: protein

A:Residues: 'X', '37, 'X', '39-45, 'X', '158-161, 'X', '163-174, 'X' <IRA>

R:Sharf, F.S.; Li, S.S.L.

J. Biol. Chem. 257, 14753-14759, 1982

A:Title: Amino acid sequence of small and large subunits of seed storage protein from

A:Reference number: A92357; MUID:83082772

A:Accession: A01328

A:Molecule type: protein

A:Residues: 157-190; 194-221, 'Q', '223-225, 230-233, 'N', '235-254, 'Q', '256-258 <SRA>

A:Note: 230-Ser was also found

A:Note: there is considerable similarity between residues 181-231 of this protein and

R:Odani, S.; Koide, T.; Ono, T.; Ohnishi, K.

Biochem. J. 213, 543-545, 1983

A:Title: Structural relationships between barley (Hordeum vulgare) trypsin inhibitor a

A:Reference number: A90322; MUID:83308577

A:Contents: annotation

A:Note: this protein is homologous with trypsin inhibitor from barley

C:Complex: consists of two chains linked by two disulfide bonds involving Cys-162 and

C:Superfamily: 2S seed storage protein precursor

C:Keywords: pyroglutamic acid; seed; storage protein
 F:1-21/Domain: signal sequence #status predicted <STIG>
 F:22-35/Domain: propeptide #status predicted <PRO>
 F:36-72/Product: probable 2S seed storage protein small chain 2 #status experimental <SC>
 F:87-156/Product: probable 2S seed storage protein large 2 #status predicted <LCHT>
 F:157-190/Product: 2S seed storage protein small chain #status experimental <SML>
 F:194-258/Product: 2S seed storage protein large chain #status experimental <LRG>
 F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 100.0%; Score 350; DB 1; Length 258;
 Best Local Similarity 100.0%; Pred. No. 1.4e-32;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QERSLRGCCDHLKQKQMSQCRCEGLRQATIEQOQSGQLGQDVFEAFRTANLPSCGVSP 60
 DB 194 QERSLRGCCDHLKQKQMSQCRCEGLRQATIEQOQSGQLGQDVFEAFRTANLPSCGVSP 253

OY 61 TECRF 65
 DB 254 TECRF 258

RESULT 2
 A59346 seed storage protein Lec2SA1 large chain [imported] - tomato (fragment)
 N:Alternate names: 2S albumin large subunit
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 31-Dec-2001 #sequence_revision 31-Dec-2001 #text_change 31-Dec-2001
 C:Accession: A59346
 R:Oguri, S.
 submitted to the Protein Sequence Database, September 2000
 A:Reference number: A59346

A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-70 <OGU>
 A:Experimental source: cv. cherry; seed
 A:Note: cross-reactive with anti-tomato lectin antiserum: one of four isoforms in seeds
 C:Complex: heterodimer of large and small (see PIR:B59346) chains, disulfide linked
 C:Keywords: heterodimer; pyroglutamic acid; seed; storage protein
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 44.4%; Score 155.5; DB 2; Length 70;
 Best Local Similarity 41.9%; Pred. No. 6.4e-11;
 Matches 26; Conservative 17; Mismatches 18; Indels 1; Gaps 1;

OY 3 RSLRGCCDHLKQKQMS-QCRCEGLRQATIEQOQSGQLGQDVFEAFRTANLPSCGVSP 61
 DB 5 RLVQCCCNQLRQKQMSQCRCEGLRQATIEQOQSGQLGQDVFEAFRTANLPSCGVSP 64

OY 62 EC 63
 DB 65 RC 66

RESULT 3
 S14947 2S albumin - Brazil nut
 C:Species: Bertholletia excelsa (Brazil nut)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S14947
 R:Gander, E.S.; Holmstrom, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross
 Plant Mol. Biol. 16, 437-448, 1991
 A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin fr
 A:Reference number: S14946; MUID:91370890
 A:Accession: S14947
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-154 <GAN>
 A:Cross-references: EMBL:X54491; NID:g17712; PIDN:CAA38363.1; PID:g17713
 C:Genetics:

A:introns: 67/3
 C:Superfamily: wheat alpha-amylase inhibitor

Query Match 42.0%; Score 147; DB 2; Length 154;
 Best Local Similarity 41.3%; Pred. No. 1.3e-09;
 Matches 26; Conservative 12; Mismatches 25; Indels 0; Gaps 0;

OY 1 QERSLRGCCDHLKQKQMSQCRCEGLRQATIEQOQSGQLGQDVFEAFRTANLPSCGVSP 60
 DB 83 EEPHLDCECCQLRMDKRCBGLRQATIEQOQSGQLGQDVFEAFRTANLPSCGVSP 142

OY 61 TEC 63
 DB 143 QRC 145

RESULT 4
 S48176 mablinin I-1 - Yunnan caper (fragments)
 C:Species: Capparis masakal (Yunnan caper)
 C:Date: 15-Jul-1995 #sequence_revision 13-Mar-1998 #text_change 13-Mar-1998
 C:Accession: S48176; S48177
 R:Nirasawa, S.; Nishino, T.; Katabira, M.; Uesugi, S.; Hu, Z.; Kurihara, Y.
 Eur. J. Biochem. 223, 989-995, 1994
 A:Title: Structures of heat-stable and unstable homologues of the sweet protein mablin
 A:Reference number: S48176; MUID:94333405
 A:Accession: S48176

A:Molecule type: protein
 A:Residues: 1-32;33-104 <NIR>
 A:Experimental source: seed
 A:Superfamily: wheat alpha-amylase inhibitor
 C:Keywords: albumin; storage protein; sweet taste
 F:1-32/Product: mablinin I-1 chain A (fragment) #status experimental <CHA>
 F:1-32,33-104/Product: mablinin I-1 (fragment) #status experimental <MAT>
 F:33-104/Product: mablinin I-1 chain B #status experimental <HB>
 F:4-53;17-42;43-91;55-99/Disulfide bonds: #status experimental

Query Match 38.6%; Score 135; DB 2; Length 104;
 Best Local Similarity 43.8%; Pred. No. 2.1e-08;
 Matches 28; Conservative 12; Mismatches 22; Indels 2; Gaps 2;

OY 4 SLRGCCDHLKQKQMSQCRCEGLRQATIEQOQSGQLGQDVFEAFRTANLPSCGVSP 61
 DB 38 ALRLCCNQLRQKQMSQCRCEGLRQATIEQOQSGQLGQDVFEAFRTANLPSCGVSP 97

OY 62 ECRF 65
 DB 98 RCOF 101

RESULT 5
 S48180 mablinin IV - Yunnan caper (fragments)
 C:Species: Capparis masakal (Yunnan caper)
 C:Date: 15-Jul-1995 #sequence_revision 13-Mar-1998 #text_change 20-Mar-1998
 C:Accession: S48180; S48181
 R:Nirasawa, S.; Nishino, T.; Katabira, M.; Uesugi, S.; Hu, Z.; Kurihara, Y.
 Eur. J. Biochem. 223, 989-995, 1994
 A:Title: Structures of heat-stable and unstable homologues of the sweet protein mablin
 A:Reference number: S48176; MUID:94333405
 A:Accession: S48180
 A:Molecule type: protein
 A:Residues: 1-28;29-100 <NIR>
 A:Experimental source: seed
 C:Keywords: albumin; storage protein; sweet taste
 F:1-28/Product: mablinin IV chain A (fragment) #status experimental <CHA>
 F:1-28,29-100/Product: mablinin IV (fragment) #status experimental <MAT>
 F:29-100/Product: mablinin IV chain B #status experimental <HB>
 F:4-49;17-38;39-87;51-95/Disulfide bonds: #status predicted

```
Query Match Summary      38.0% Score 133; DB 2; Length 100;
Best Local Similarity   43.8%; Pred. No. 3.4e-06;
Matches    28; Conservative    11; Mismatches    23; Indels     2; Gaps     2;
```

OY 4 SLRGCCDHLKQMQSQRCGEGLRQAIEQQQSOGLOG-OVDEAFRTAANLEPSCVSPT - 61
 :|||:::||:-|-|-----|||----|||---:|:||||:|
Db 34 ALRLCCNLRQYNKPCCVCPLVQLAAHQOLLYQGILGPGVRRLFRANLNPICKIIPAVG 93
 :|||:|:|--|||

OY 62 ECRF 65
 ||:|

Db 94 RCQF 97

RESULT 6
S48178
 mabnl1n III - Yunnan caper (fragments)
C:Species: Caparis masakaal (Yunnan caper)
C:date: 15-Jul-1995 #sequence_revision 13-Mar-1998 #text_change 20-Mar-1998
C:Accession: S48178; S48179
R:Nirasawa, S.; Nishino, T.; Katahira, M.; Uesugi, S.; Hu, Z.; Kurihara, Y.
E:J. Blochem. 223, 989-995, 1994
A:title Structures of heat-stable and unstable homologues of the sweet protein mabnl1n
A:Reference number: S48176; MUID:94333405
A:Accession: S48178
A:Molecule type: protein
A:Residues: 1-32,33-104 <NIR>
A:Experimental source: seed
C:Superfamily: wheat alpha-amylase inhibitor
C:keywords: albumin storage protein; sweet taste
F:1-32/Product: mabnl1n III chain A (fragment) #status experimental <CHA>
F:1-32,33-104/Product: mabnl1n III (fragment) #status experimental <CHB>
F:33-104/Product: mabnl1n III chain B #status experimental <MAT>
F:4-53,17-42,43-91,55-99/Disulfide bonds: #status predicted

A:Cross-references: EMBL:X57027; NID:g17714; PID:g17715
R:Altenbach, S.B.; Pearson, K.W.; Leung, F.W.; Sun, S.S.M.
Plant Mol. Biol. 8, 239-250, 1987
A>Title: Cloning and sequence analysis of a cDNA encoding a Brazil nut protein except
A:Reference number: S06252
A:Accession: S06252
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-146 <ALT>
A:Experimental source: clone pHS-3
A>Note: part of this sequence, including the amino end of the large chain, was confir
R:Bassuener, R.; Schlesier, B.
submitted to the EMBL Data Library, December 1990
A:Reference number: S21640
A:Accession: S21640
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <BAZ>
A:Cross-references: EMBL:X57028; NID:g17716; PID:g17717
R:Ampe, C.; Van Damme, J.; de Castro, L.A.B.; Sampelo, M.J.A.M.; Van Montagu, M.; Van
Eur. J. Biochem. 159, 597-604, 1986
A>Title: The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil
A:Reference number: A91173; MUID:87004679
A:Accession: B25802
A:Molecule type: protein
A:Residues: 70-90, 'E', 92-121, 'W', 123-125, 'L', 127-142 <AMP>
C:Genetics:
A:Introns: 60/3
C:Superfamily: Wheat alpha-amylase inhibitor
F:1-72/Domain: signal sequence #status predicted <IG>
F:23-46/Domain: propeptide #status predicted <PRO>
F:47-69/Product: seed storage protein small chain #status predicted <SCH>
F:70-146/Product: seed storage protein large chain #status experimental <LCH>

Query Match	38.0%;	Score 133;	DB 2;	Length 104;	
Best Local Similarity	43.8%;	Pred. No. 3.5e-08;			
Matches	28;	Conservative 11;	Mismatches 23;	Indels 2;	Gaps 2.
QY	4	SLRGCCDLKOMOSQRCCEGLNQAIEQQDSQGLQG-QDVEAFRTAALPSMCGVSP	-	61	
		:: ::			::
Db	38	ALRLCCNQLDQVNRKCVQVRLNQAHHQDLQGLDQISPRQVRLPLFRARLPLNICKIPAVG		97	
QY	62	ECRF	65		
		::			
Db	98	RCQF	101		

```

Query Match          37.9%  Score 132.5.  DB 2;  Length 146;
Best Local Similarity 39.7%;  Pred. No. 5.6e-08;
Matches 25;  Conservative 10;  Mismatches 27;  Indels 1;  Gaps 1;

OY 2 ERLSGCDHLKQOMSGCCGGLR-QAIEQQSGGQLGSDVFEAFRTANLPSMGVSP 60
   | : ||: | : | ||||| : || : : : | : | : | : | : | : ||
Db 75 EPHNSECCEQLGMDSESCRGELRMHMMHMOQEEHQPGEDQRRMRLAENIPSCNLS 134

OY 61 TEC 63
   |
Db 135 MRC 137

```

```

RESULT      7
SI4946      2S seed storage protein large chain - Brazil nut
N:Alternate names: albumin 2S precursor
C:Species: Bertholletia excelsa (Brazil nut)
C:Date: 21-Nov-1993 #sequence.revision 26-May-1995 #text.change 21-Jul-2000
C:Accession: SI4946; SI4479; S06252; S21640; B25802
R:Gander, E.S.; Holmstrom, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross
Plant Mol. Biol. 16, 437-448, 1991
A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin fr
A:Reference number: SI4946; MUID:91370890
A:Accession: SI4946
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <GAN>
A:Cross-references: EMBL:X54490; NID:g17710; PIDN:CAA38362.1; PID:g17711
A:Note: the authors translated the codon CTT for residue 13 as Val and GTC for residue 1
A:Bassener, R.; Schlesier, B.
submitted to the EMBL data library, December 1990
A:Reference number: SI4479
A:Accession: SI4479
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <BAS>

```

RESULT 8
T10257
2S albumin precursor - cucurbit
N:Alternate names: prepro2S albumin
C:Species: Cucurbita sp. (cucurbit)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10257; S19323
R:Hara-Nishimura, I.; Takeuchi, Y.; Inoue, K.; Nishimura, M.
A:Title: Vesicle transport and processing of the precursor to 2S albumin in pumpkin.
A:Reference number: 217000; MUID:94100993
A:Accession: T10257
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA; protein
A:Residues: 1-141 <HAR1>
A:Cross-references: EMBL:DL1560; NID:9459404; PIDN:BA03993.1; PID:9459405
A:Experimental source: seed, storage deposition stage; cotyledon
A>Note: soluble seed protein
R:Hara-Nishimura, I.; Inoue, K.; Nishimura, M.
FEBS Lett. 294, 89-93, 1991
A:Title: A unique vacuolar processing enzyme responsible for conversion of several pro
A:Reference number: S19323; MUID:9207151
A:Accession: S19323
A:Status: preliminary
A:Molecule type: mRNA

Db 73 CCQOLEKMDTQCRCQGLRHATMQOMQOMGMSKQREIMQVTKRKIMSECEMEPEGRG 131

RESULT 13

NRMT1

2S albumin 1 precursor - Arabidopsis thaliana
N:Alternate names: seed storage protein AT2S1

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Dec-1999

C:Accession: JAO161: PS0282; S34676: T06044

R:Kreberg, E.; Hertles, L.; de Clercq, A.; Seurinck, J.; Leemans, J.; Van Damme, J.; Se

Plant Physiol. 87, 859-866, 1988

A>Title: Determination of the processing sites of an Arabidopsis 2S albumin and character

A:Reference number: JAO161

A:Accession: JAO161

A:Molecule type: DNA

A:Residues: 1-164 <KRE>

A:Cross-references: GB:M22032; NID:g166609; PIDN:AAA32743.1; PID:g166614

A:Accession: PS0282

A:Molecule type: protein

A:Residues: 38-73:84-162 <KR2>

R:Conceicao, A.D.S.; Kreberg, E.

submitted to the EMBL Data Library, July 1993

A:Description: Tentative title: a cotyledon regulatory region is responsible for the dif

A:Reference number: S34674

A:Accession: S34676

A:Molecule type: DNA

A:Residues: 1-164 <CON>

A:Cross-references: EMBL:224745; NID:g395203; PIDN:CAA80870.1; PID:g395204

R:Hevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15484

A:Accession: T06044

A:Molecule type: DNA

A:Residues: 1-164 <BEV>

A:Cross-references: EMBL:AL035680; GSPDB:GN00062; ATSP:T24A18.90

A:Experimental source: cultivar Columbia; BAC clone T24A18

C:Genetics:

A:Gene: T24A18.90

A:Map position: 4

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: seed; storage protein

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-164/Product: 2S albumin 1 proprotein #status predicted <AN2>

F:38-73/Product: 2S albumin 1 small chain #status experimental <SMC>

F:84-162/Product: 2S albumin 1 large chain #status experimental <LAC>

Query Match 32.0%; Score 112; DB 1; Length 164;

Best Local Similarity 35.8%; Pred. No. 1.4e-05;

Matches 24; Conservative 16; Mismatches 25; Indels 2; Gaps 2;

OY 1 QENSLSRGCCDHLKQMSQCRCEGLRQALIEQOOSQGLQGVDFEAFRTANLPSCGVS 59

Db 89 QEOQLFOCCNEIRQEBPCVPTLKQAKAVRLOGHOPQYRKIKYQTAKEHPNVCIDIP 148

OY 60 PTE-CRF 65

Db 149 QVDVCPF 155

RESULT 14

S38887

2S albumin - common sunflower (fragment)

C:Species: Helianthus annuus (common sunflower)

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-May-2000

C:Accession: S38887

R:Thoyts, P.J.E.; Millchip, M.; Stobart, A.K.; Griffiths, W.T.; Napier, J.A.; Shewry, F

submitted to the EMBL Data Library, November 1993

A:Reference number: S38887

A:Accession: S38887

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-323 <THO>
A:Cross-references: EMBL:X76101; NID:g429181; PID:g429182
C:Superfamily: gliadin

Query Match 31.6%; Score 110.5; DB 2; Length 323;

Best Local Similarity 30.9%; Pred. No. 4e-05;

Matches 21; Conservative 15; Mismatches 27; Indels 5; Gaps 1;

OY 1 QENSRLCCCHLKQMSQCRCE-----GLQALIEQOOSQGLQGVDFEAFRTANLPSCM 55

Db 252 EQQGLQCCNEILQVNRQCCALKEVGRMRQOQOQRRQYGGQYQYVERILENLPNQ 311

OY 56 CGVSPTFC 63

Db 312 CDLDVQC 319

RESULT 15

S46514

putridoline-b precursor - wheat

C:Species: Triticum aestivum (common wheat)

C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Aug-1999

C:Accession: S46514

R:Gautier, M.F.; Aleman, M.E.; Guirao, A.; Marion, D.; Joudrier, P.

Plant Mol. Biol. 25, 43-57, 1994

A>Title: Triticum aestivum putridolines, two basic cysteine-rich seed proteins: cDNA

A:Reference number: S46514; M01D:94272013

A:Accession: S46514

A:Molecule type: mRNA

A:Residues: 1-148 <GAU>

A:Cross-references: EMBL:X69912; NID:g509087; PIDN:CAA49537.1; PID:g509088

C:Superfamily: glycine

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-145/Product: putridoline-a #status predicted <MAT>

F:146-148/Domain: carboxyl-terminal propeptide #status predicted <PRO>

Query Match 30.9%; Score 108; DB 2; Length 148;

Best Local Similarity 35.9%; Pred. No. 3.6e-05;

Matches 23; Conservative 12; Mismatches 17; Indels 12; Gaps 3;

OY 8 CCHLQMSQCRCEGLRQALIEQOOSQGLQGVDFEAFRTANLPSCGVSPT 61

Db 85 CCKQLSGLAPQCRCDISIRVI-----QGRUGFLGIRGEVYQLODRAQSLPSKCMNG-A 138

OY 62 ECRF 65

Db 139 DCKF 142

Search completed: September 9, 2002, 12:43:31
Job time: 223 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:46:06 ; Search time 22.97 Seconds
(without alignments)
109.568 Million cell updates/sec

Title: US-09-913-351-5
Perfect score: 350
Sequence: 1 QERSLRCCCHIKQMOSQR.....FRFANLPSCMGVPTRECRF 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	350	100.0	258	1	P01089 ricinus com
2	350	100.0	258	1	P01089 ricinus com
3	133	38.0	100	1	P80351 caparis ma
4	133	38.0	100	1	P80351 caparis ma
5	132.5	37.9	146	1	P04403 beithollet
6	113.5	34.1	295	1	P13461 helianthus
7	112	32.0	164	1	P13457 arabidopsis
8	108	30.9	148	1	P01090 brassica na
9	106.5	30.4	155	1	P01090 brassica na
10	104.5	29.9	178	1	P26986 picea glauc
11	104	29.7	172	1	P27740 brassica na
12	102.5	29.3	178	1	P27740 brassica na
13	102	29.1	130	1	P38057 sinapis arv
14	100.5	28.7	129	1	P80207 brassica ju
15	98.5	28.1	186	1	P09693 brassica na
16	96.5	27.6	180	1	P17333 brassica na
17	96.5	27.6	180	1	P08453 triticum ae
18	95	27.1	166	1	P13460 arabidopsis
19	94.5	27.0	170	1	P13460 arabidopsis
20	92	26.3	80	1	P09693 brassica na
21	91.5	26.1	186	1	P09693 brassica na
22	89.5	25.6	60	1	P24033 moringa ole
23	89.5	25.6	60	1	P24033 moringa ole
24	88.5	25.3	127	1	P15322 sinapis alb
25	88	25.1	123	1	P15322 sinapis alb
26	86.5	24.7	133	1	P04706 zea mays (m
27	86	24.6	144	1	P01091 lupinus ang
28	84.5	24.1	149	1	P01086 hordeum vul
29	83.5	23.9	148	1	P20698 oryza sativ
30	82.5	23.6	305	1	P33432 triticum ae
31	82	23.4	110	1	P17990 hordeum vul
32	80	22.9	123	1	P24965 brassica na
33	79	22.6	286	1	P01083 triticum ae

34	79	22.6	356	1	GLTA_WHEAT	P10385 triticum ae
35	75	21.4	307	1	GDA9_WHEAT	P18573 triticum ae
36	75	21.3	307	1	GLTB_WHEAT	P10386 triticum ae
37	74.5	21.3	158	1	ZSS_SOYBN	P19594 glycine max
38	73.5	21.0	117	1	NR5_ORYSA	O65091 oryza sativ
39	73.5	21.0	296	1	GDA6_WHEAT	P04726 triticum ae
40	73.5	21.0	297	1	GDA4_WHEAT	P04724 triticum ae
41	73	20.9	124	1	IRAI_WHEAT	P01085 triticum ae
42	73	20.9	214	1	AVEN_AVEA	P27919 avena sativ
43	73	20.9	291	1	GDBH_WHEAT	P06659 triticum ae
44	72	20.6	156	1	PRO2_ORYSA	P17048 oryza sativ
45	72	20.6	244	1	GDB3_WHEAT	P04730 triticum ae

ALIGNMENTS

RESULT 1
ID 2SS_RICCO STANDARD: PRT; 258 AA.
AC P01089; 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S albumin precursor (Allergen Ric c 1).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosid 1; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91109729; PubMed=2274038;
RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor: a single preproprotein
Mol. Gen. Genet. 222:400-408(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91016940; PubMed=2216785;
RA Irwin S.D., Lord J.M.;
RT "Nucleotide sequence of a Ricinus communis 2S albumin precursor
gene.";
RL Nucleic Acids Res. 18:5890-5890(1990).
RN [3]
RP SEQUENCE OF 157-190 AND 194-258.
RX MEDLINE=83082772; PubMed=7174664;
RA Sharief F.S., Li S.S.-L.;
RT "Amino acid sequence of small and large subunits of seed storage
protein from Ricinus communis.";
RL J. Biol. Chem. 257:14753-14759(1982).
RN [4]
RP SIMILARITY TO PROTEINASE INHIBITORS.
RX MEDLINE=83308577; PubMed=6515448;
RA Odani S., Koide T., Ono T., Ohnishi K.;
RT "Structural relationship between barley (Hordeum vulgare) trypsin
inhibitor and castor-bean (Ricinus communis) storage protein.";
RL Biochem. J. 213:543-545(1983).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
LINKED BY 2 DISULFIDE BONDS.
CC -1- PTM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE
CHAINS INVOLVE CYS-162 AND CYS-175.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@sib.ch.

CC -----
DR EMBL: X54158; CAA38097.1; -
DR PIR: A01328; RZCS.
DR PIR: S11499; S11499.
DR PIR: S11500; S11500.
DR PIR: S11501; S11501.
DR InterPro: IPR003612; AAI.
DR Pfam: PF00234; trypt_alpha-amyl_2.
DR SMART: SM00499; AAI; 2.
KW Seed storage protein; Signal; Allergen.
FT SIGNAL 1
FT PROPEP 22 156 2S ALBUMIN, SMALL CHAIN.
FT CHAIN 157 190
FT PROPEP 191 193
FT CHAIN 194 258 2S ALBUMIN, LARGE CHAIN.
FT MOD_RES 194 194 PYRROLIDONE CARBOXYLIC ACID.
FT CONFLICT 222 222 E -> Q (IN REF. 3).
FT CONFLICT 226 229 MISSING (IN REF. 3).
FT CONFLICT 234 234 D -> N (IN REF. 3).
FT CONFLICT 255 255 E -> Q (IN REF. 3).
SQ SEQUENCE 258 AA; 29290 MW; 27874CFC50E41072 CRC64;

Query Match 100.0%; Score 350; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 5.8e-33;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QSRJRGCCDHKQMOQSCRCGLRQATEEQOQOGQDVFQAFRTAANLPSMCGVSP 60
DB 194 QSRJRGCCDHKQMOQSCRCGLRQATEEQOQOGQDVFQAFRTAANLPSMCGVSP 253
OY 61 TECRF 65
DB 254 TECRF 258

RESULT 2
ID 2SS1_CAPMA STANDARD; PRT; 104 AA.
AC P80351.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin I-1, A and B chains (MAB I) (Sweet protein).
OS Capparis masakal (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Capparidaceae;
OX NCBI_TaxID=13395;
RN [1]
RP TISSUE=Seed;
RC MEDLINE=94333405; PubMed=8055976;
RA Nirasawa S., Nishino T., Katahira M., Uesugi S., Hu Z., Kurihara Y.;
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.";
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: 2S SEED STORAGE PROTEIN HAVING SWEETNESS-INDUCING
CC ACTIVITY. THIS FORM IS NOT HEAT STABLE.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 32 MABINLIN I, A CHAIN.

FT NON_CONS 32 33
FT CHAIN 33 104 MABINLIN I, B CHAIN.
FT DISULFID 4 53
FT DISULFID 17 42
FT DISULFID 43 91
FT DISULFID 55 99
SQ SEQUENCE 104 AA; 12284 MW; 156A4A4B50D5EE60 CRC64;

Query Match 38.6%; Score 135; DB 1; Length 104;
Best Local Similarity 43.8%; Pred. No. 7.5e-09;
Matches 28; Conservative 12; Mismatches 22; Indels 2; Gaps 2;

OY 4 SIRGCCDHKQMOQSCRCGLRQATEEQOQOGQDVFQAFRTAANLPSMCGVSP - 61
DB 38 ALRCCNOLROYNKPVCVVLQAAHQLYQGIIEBPRQVRLFRARNLPNICKIPAVG 97
OY 62 ECRF 65
DB 98 RCOF 101

RESULT 3
ID 2SS4_CAPMA STANDARD; PRT; 100 AA.
AC P80353;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin IV, A and B chains (MAB IV) (Sweet protein).
OS Capparis masakal (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Capparidaceae;
OX NCBI_TaxID=13395;
RN [1]
RP TISSUE=Seed;
RC MEDLINE=94333405; PubMed=8055976;
RA Nirasawa S., Nishino T., Katahira M., Uesugi S., Hu Z., Kurihara Y.;
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.";
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
CC INDUCING ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 28 MABINLIN IV, A CHAIN.
FT NON_CONS 28 29
FT CHAIN 29 100 MABINLIN IV, B CHAIN.
FT DISULFID 4 49 BY SIMILARITY.
FT DISULFID 17 38 BY SIMILARITY.
FT DISULFID 39 87 BY SIMILARITY.
FT DISULFID 51 95 BY SIMILARITY.
SQ SEQUENCE 100 AA; 11928 MW; 7B09673FAB7793CA CRC64;

Query Match 38.0%; Score 133; DB 1; Length 100;
Best Local Similarity 43.8%; Pred. No. 1.2e-08;
Matches 28; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

OY 4 SIRGCCDHKQMOQSCRCGLRQATEEQOQOGQDVFQAFRTAANLPSMCGVSP - 61
DB 34 ALRCCNOLROYNKPVCVVLQAAHQOOLYQGIIEBPRQVRLFRARNLPNICKIPAVG 93

QY 62 ECRF 65
 Db 94 RQCF 97

RESULT 4
 2SS3_CAPMA STANDARD; PRT; 104 AA.
 ID 2SS3_CAPMA
 AC P80352;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mabinlin III, A and B chains (MAB III) (Sweet protein).
 OS Capparis masalai (Mabinlin).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Caparis.
 OX NCBI_TaxID=13395;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Seed;
 RX MEDLINE=94333405; PubMed=8055976;
 RA Nirasawa S., Nishino T., Katakura M., Uesugi S., Hu Z., Kurihara Y.;
 RT "Structures of heat-stable and unstable homologues of the sweet
 RT protein mabinlin. The difference in the heat stability is due to
 RT replacement of a single amino acid residue."
 RL Eur. J. Biochem. 223:989-995(1994).
 CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
 CC INDUCING ACTIVITY.
 CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
 CC DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF01631; Seedstora_2s; 1.
 DR PRINTS: PR00496; NAPIIN.
 DR ProDom: PD002498; Napin; 1.
 DR SMART: SM00499; AAI; 1.
 KW Seed storage protein; Albumin; Sweet-taste.
 FT CHAIN 1 32 MABINLIN III, A CHAIN.
 FT NON_CONS 32 33
 FT CHAIN 33 104 MABINLIN III, B CHAIN.
 FT DISULFID 4 53
 FT DISULFID 17 42
 FT DISULFID 43 91
 FT DISULFID 55 99
 SQ SEQUENCE 104 AA; 12284 MW; 102BE3F5F24AD3D0 CRC64;

Query Match 38.0%; Score 133; DB 1; Length 104;
 Best Local Similarity 43.8%; Pred. No. 1.3e-08;
 Matches 28; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 4 SLRGCCDHLKQMSQCEGLROAIEQOQSGOLGQDVEFAFRTPANLPSMGVSPPT- 61
 Db 38 ALRLCCNQLKQVKNPCVPLRQAHQQLYQGLIEGRVRLFRARNLPICKIPAVG 97

QY 62 ECRF 65
 Db 98 RQCF 101

RESULT 5
 2SS_BEREX STANDARD; PRT; 146 AA.
 ID 2SS_BEREX
 AC P04403; P04402;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 2S sulfur-rich seed storage protein precursor (Allergen Ber e 1).
 GN BE2S1 AND BE2S2.
 OS Bertholletia excelsa (Brazil nut).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Ericales; Lecythidaceae; Bertholletia.
 OX NCBI_TaxID=3645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M.;
 RT "Cloning and sequence analysis of a cDNA encoding a Brazil nut protein
 RT exceptionally rich in methionine."
 RL Plant Mol. Biol. 8:239-250(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bassener R.;
 RL Submitted (DEC-1990) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91370890; PubMed=1840683;
 RA Gander E.S., Holmstroem K.O., de Paiva G.R., de Castro L.A.B.,
 RA Carneiro M., Grossi de Sa M.F.;
 RT "Isolation, characterization and expression of a gene coding for a 2S
 RT albumin from Bertholletia excelsa (Brazil nut)."
 RL Plant Mol. Biol. 16:437-448(1991).
 RN [4]
 RP SEQUENCE OF 37-64 AND 70-142.
 RX MEDLINE=87004679; PubMed=3758080;
 RA Ampe C., van Damme J., de Castro L.A.B., Sampaio M.J.A.M.,
 RA van Montagu M., Vandereckhove J.;
 RT "The amino-acid sequence of the 2S sulphur-rich proteins from seeds
 RT of Brazil nut (Bertholletia excelsa H.B.K.)."
 RL Eur. J. Biochem. 159:597-604(1986).
 CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M17146; AAA33010.1; -;
 DR EMBL: X57027; CAA40343.1; -;
 DR EMBL: X57028; CAA40344.1; -;
 DR EMBL: X54490; CAA38362.1; -;
 DR EMBL: X54491; CAA38363.1; ALT_SEQ.
 DR EMBL: A13818; CAA01131.1; -;
 DR PIR: S06252; S06252.
 DR PIR: A25802; A25802.
 DR PIR: B25802; B25802.
 DR PIR: S14946; S14946.
 DR PIR: S14479; S14479.
 DR PIR: S21640; S21640.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001766; Cereals_tryp_amyl_inh.
 DR Pfam: PF00234; tryp_alpha_amyl; 1.
 DR SMART: SM00499; AAI; 1.
 KW Seed storage protein; Signal; Allergen.
 FT SIGNAL 1 22
 FT PROPEP 23 36
 FT CHAIN 37 64 SMALL CHAIN.
 FT PROPEP 65 69
 FT CHAIN 70 142 LARGE CHAIN 1B.
 FT PROPEP 143 146
 FT MOD_RES 37 37
 FT VARIANT 91 91
 FT CONFLICT 38 39 S->E (IN VARIANT 1A).
 FT CONFLICT 122 122 EE->Q (IN REF. 4).
 FT CONFLICT 126 126 L->M (IN REF. 4).
 FT CONFLICT 126 126 I->L (IN REF. 4).
 SQ SEQUENCE 146 AA; 16911 MW; A7DF778BD766410D CRC64;

Query Match 37.9%; Score 132.5; DB 1; Length 146;
Best Local Similarity 39.7%; Pred. No. 2e-08;
Matches 25; Conservative 10; Mismatches 27; Indels 1; Gaps 1;

OY 2 ERSLRCCDHLKOMQSCRCRGLR-0AIEQSQSGQLQGDVEAFRTAANLPSMCGVSP 60
DB 75 EPMMSCECEQLEGMDESCRCEGLRMMRMQCEMOPRGROMRRMLAENIPSRCLTSP 134
OY 61 TEC 63
DB 135 MRC 137

RESULT 6
ID 2SS5_HELAN STANDARD; PRT; 295 AA.
AC P15461;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 2S seed storage protein precursor (2S albumin storage protein).
GN HAG5.
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids II; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 162-173.
RC STRAIN=CV. GIANT GREY STRIPE;
RX MEDLINE=88142538; PubMed=2830455;
RA Allen R.D., Cohen E.A., Vander Haar R.A., Adams C.A., Ma D.P.,
RA Nessler C.L., Thomas T.L.,
RT "Sequence and expression of a gene encoding an albumin storage
RT protein in sunflower."
RL Mol. Genet. 210:211-218(1987).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- PFM: IT IS POSSIBLE THAT THE 38 KDA PRECURSOR IS CLEAVED INTO TWO
CC POLYPEPTIDES THAT ARE APPROXIMATELY THE SAME SIZE. THE MATURE
CC PROTEIN IS COMPOSED OF A SINGLE POLYPEPTIDE CONTAINING ONE OR MORE
CC INTRA-MOLECULAR DISULFIDE LINKAGES.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: X06410; CAA29699.1; -
DR PIR: S01062; S01062.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereals-tryp_aml1_inh.
DR Pfam: PF00234; tryp_alpha_aml1_2.
DR SMART: SM00499; AAI; 2.
KW Seed storage protein; Albumin; Signal; Multigene family.
FT SIGNAL 1 20
FT PROPEP 21 161 POTENTIAL.
FT CHAIN 162 295 2S SEED STORAGE PROTEIN.
FT SEQUENCE 295 AA; 34071 MW; 8958A106805142A1 CRC64;

Query Match 34.1%; Score 119.5; DB 1; Length 295;
Best Local Similarity 30.3%; Pred. No. 1.2e-06;
Matches 23; Conservative 17; Mismatches 23; Indels 13; Gaps 2;

OY 1 OERSLUG-----CCDHLKOMQSCRCRGLR0AI-----EEQSQSGQLQGDVEAFR 47
DB 215 QQQQQQRLQQQCCNQLQNVKRECHCAIDQVARRVRKRFQQQQQQRRGGFGGQGMETARR 274

OY 48 TAANLPSMCGVSPTTEC 63
DB 275 VIONLPRQCDELEVOQC 290

RESULT 7
ID 2SSL_ARATH STANDARD; PRT; 164 AA.
AC P15457;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S seed storage protein 1 precursor (2S albumin storage protein)
DE (NMW02-2S albumin 1).
GN AT2S1 OR AT4G27140 OR T2A18.90.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-73 AND 84-162.
RC STRAIN=CV. C24;
RA Kriebbers E., Herdies L., de Clercq A., Seurinck J., Leemans J.,
RA Van Damme J., Segura M., Gheysen G., van Montagu M.,
RA Vandekerckhove J.;
RT "Determination of the processing sites of an Arabidopsis 2S albumin
RT and characterization of the complete gene family."
RL Plant Physiol. 87:859-866(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RA Conceicao A.D.S., Kriebbers E.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Scheller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entlin K.-D., Terry N.,
RA Harris B., Ansoorge W., Brandt P., Griwell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermayer B., Maché R., Mueller M.,
RA Kreis M., Delseny M., Puldomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portelle D., Perez-Alonso M., Botry M., Bancroft I.,
RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Welljens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Holtelegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirksen W.,
RA Moollman P., Klein lankhorst R., Rose M., Hauf J., Koetler P.,
RA Bernelsen S., Hempel S., Feldpausch M., Lambertz S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Gronin A., Quail M., Bryl-Alen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McWay K., Mayes R.,
RA Petrelet A., Rajandream M.A., Lyne M., Benes V., Rechner S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Danner D., Herzl A.,
RA Neumann S., Argitron A., Vitale D., Liguori R., Piravandi E.,
RA Massenet O., Quidley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Cheifor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bietke C.,
RA Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Clout J., Abbott A., Scott K., Johnson D.,

Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 Kramar J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 Nelson J., Speth J., Ryan E., Andrews S., Geisel C., Layman D.,
 Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 Ma P., Zhong J., Preston R., Vill D., Shekher M., Mateo A., Shah R.,
 Swamy I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Gratz S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Marienssen R., McCombie W.R.;
 "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY DISULFIDE BONDS.
 CC -1- MISCELLANEOUS: THIS IS THE MOST ABUNDANT ISOFORM OF 2S ALBUMIN IN
 CC ARABIDOPSIS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M22032; AAA32743.1; -;
 DR EMBL: Z24745; CAA80870.1; -;
 DR EMBL: A13820; CAA01132.1; -;
 DR EMBL: AL035680; CAB38844.1; -;
 DR EMBL: AL161566; CAB79569.1; -;
 DR PIR: JAO161; NMM01.
 DR PIR: S34676; S34676.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal-tryp_amy1_inh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_amy1.1.
 DR PRINTS: PR00496; NAPIN.
 DR PRODOM: PD002498; Napin.1.
 DR SMART: SM00499; AAI.1.
 KW Seed storage protein; Albumin; signal; Multigene family.
 FT SIGNAL 1 21
 FT PROPEP 22 37
 FT CHAIN 38 73 2S SEED STORAGE PROTEIN 1 SMALL SUBUNIT.
 FT PROPEP 74 83
 FT CHAIN 84 162 2S SEED STORAGE PROTEIN 1 LARGE SUBUNIT.
 FT PROPEP 163 164
 SQ SEQUENCE 164 AA; 19014 MW; 2BF28CB474D9832B CRC64;

Query Match 32.0%; Score 112; DB 1; Length 164;
 Best Local Similarity 35.8%; Pred. No. 4.9e-06;
 Matches 24; Conservative 16; Mismatches 25; Indels 2; Gaps 2;
 Oy 1 GERSL-RGCGCHLKQMSQRCRGRLQALIEQOOSGQAGQGVFAFRNALPSCGVS 59
 Db 89 QEOQLPQCCNCELRQEDPCVCTLKQAKAVRLOGHQHPQVRIYGTAKHLPNVCIP 148
 Oy 60 PTE-CRF 65
 Db 149 QVDVCPF 155

RESULT 8
 ID PUTB_WHEAT STANDARD; PRT; 148 AA.
 AC 010464;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Puroindoline-B precursor.
 OS Triticum aestivum (wheat).
 OS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CAPITOLE; TISSUE=Seed;
 RX MEDLINE=94272013; PubMed=7516201;
 RA Gautier M.-F., Aleman M.-F., Guirao A., Marion D., Joudrier P.;
 RT "Triticum aestivum puroindolines, two basic cysteine-rich seed
 RT proteins: cDNA sequence analysis and developmental gene expression.";
 RL Plant Mol. Biol. 25:43-57(1994).
 RN [2]
 RP SEQUENCE OF 30-148.
 RA Blochet J.E., Kaboulou A., Compont J.P., Marion D.;
 RL (in) Bushuk W., Tkachuk R. (eds.);
 RL Gluten proteins, pp.314-325, American Association of Cereal Chemists,
 RL St. Paul MI (1991).
 CC -1- FUNCTION: ACTS AS A MEMBRANOTOXIN, PROBABLY THROUGH ITS
 CC ANTIMICROBIAL AND ANTIFUNGAL ACTIVITIES, CONTRIBUTING TO THE
 CC DEFENSE MECHANISM OF THE PLANT AGAINST PREDATORS.
 CC -1- PTM: FIVE DISULFIDE BONDS ARE PRESENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X69912; CAA49537.1; -;
 DR HSSP: P07597; IITP.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal-tryp_amy1_inh.
 DR Pfam: PF00234; tryp_alpha_amy1.1.
 DR SMART: SM00499; AAI.1.
 KW Membrane; Toxin; Antibiotic; Signal.
 FT SIGNAL 1 19
 FT PROPEP 20 29 PUROINDOLINE-B.
 FT CHAIN 30 148 TRP-RICH.
 FT DOMAIN 68 73
 SQ SEQUENCE 148 AA; 16792 MW; 327904B4BEBC2C16 CRC64;

Query Match 30.9%; Score 108; DB 1; Length 148;
 Best Local Similarity 35.9%; Pred. No. 1.3e-05;
 Matches 23; Conservative 12; Mismatches 17; Indels 12; Gaps 3;
 Oy 8 CCDHLKQMSQRCRGRLQALIEQOOSGQAGQGVFAFRNALPSCGVSPT 61
 Db 85 CCKOLSOIAPQCRDCSIRRVI-----QGRIGGFLGIRGEVFKOLRAQSLPSKCMNG-A 138
 Oy 62 ECRF 65
 Db 139 DCKF 142

RESULT 9
 ID 2SS2_CAPMA STANDARD; PRT; 155 AA.
 AC P30233; C04774;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Madinilla II precursor (MAB II) (sweet protein).
 OS Capparis masakal (Mabianlang).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Capparis.
 OX NCBI_TaxID=13395;
 RN [1]
 RP SEQUENCE FROM N.A.

Query Match	Best Local Similarity	Score	DB 1;	Length	155;
Matches	22;	Conservative	11;	Mismatches	22;
				Indels	1;
				Gaps	1;
QY	4	SLRGCCDHLKQMSQRCGELRQATEQQSQGLQG-ODVFEAFRTAANLPSMGCV	58		
Db	88	ALROCCNOLROVDRCVCVPVLRQAQOVLQROIIIGPOOLRLFLDARLPLNICNI	143		
RESULT	10				
ID	2552_BRANA	STANDARD;	PRT;	178	AA.

```

AC      P01090; 1986 (Rel. 01, Created)
DT      01-JUL-1986 (Rel. 01, Last sequence update)
DT      01-MAR-1989 (Rel. 10, Last annotation update)
DT      01-AUG-1992 (Rel. 23, Last annotation update)
DE      Napin 2 precursor (1.7S seed storage protein).
OS      Brassica napus (Rape).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC      Spermatophytes; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Brassica.
CX      NCBI_TaxID=3708;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=87308224; Pubmed=3624251;
RA      Josefsson L.-G., Lemman M., Ericson M.L., Rask L.;
RT      "Structure of a gene encoding the 1.7 S storage protein, napin, from
RL      Brassica napus."
RL      J. Biol. Chem. 262:12196-12201(1987).
RN      [2]
RP      REVISIONS.
RA      Josefsson L.-G.;
RL      Submitted (JUL-1987) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=87033665; Pubmed=3771543;
RA      Ericson M.L., Koedlin J., Lemman M., Glimelius L.-G.,
RL      Rask L.;
RT      "Structure of the rapeseed 1.7 S storage protein, napin, and its
RT      precursor."
RL      J. Biol. Chem. 261:14576-14581(1986).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV, TOWER;
RX      MEDLINE=84113267; Pubmed=6689334;
RA      Crouch M.L., Tenbarghe K.M., Simon A.E., Ferl R.;
RT      "cDNA clones for Brassica napus seed storage proteins: evidence from
RT      nucleotide sequence analysis that both subunits of napin are cleaved
RT      from a precursor polypeptide."
RL      J. Mol. Appl. Genet. 2:273-283(1983).
CC      -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
CC      TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC      ITS MATURATION.
CC      -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC      LINKED BY DISULFIDE BONDS.
CC      -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
CC      -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; K01545; AAA33006.1; -
DR      EMBL; J02586; AAA32997.1; -
DR      EMBL; J02798; AAA87348.1; -
DR      PIR; A01329; NMRP2.
DR      PIR; A29801; A29801.
DR      PIR; A25997; A25997.
DR      InterPro; IPR003612; AAI.
DR      InterPro; IPR001768; Cereal_tryp_amyl_1nh.
DR      InterPro; IPR000617; Napin.
DR      Pfam; PF00234; tryp_alpha_amyl_1.
DR      PRINTS; PR00496; NAPIN.
DR      PRODOM; PD002498; Napin; 1.
DR      SMART; SM00499; AAI; 1.
KW      Seed storage protein; Signal; Multigene family.
FT      SIGNAL 1 21
FT      PROPEP 22 38
FT      CHAIN 39 74 SMALL CHAIN.
FT      PROPEP 75 94
FT      CHAIN 95 175 LARGE CHAIN.

```


FT CONFLICT 37 37 D -> N (IN REF. 4).
 FT CONFLICT 76 76 S -> N (IN REF. 4).
 SQ SEQUENCE 178 AA; 20104 MW; 734E561971B539FF CRC64;

Query Match 29.9%; Score 104.5; DB 1; Length 178;
 Best Local Similarity 38.2%; Pred. No. 3.8e-05;

Matches 26; Conservative 12; Mismatches 23; Indels 7; Gaps 4;

OY 5 LRGCCDHLKOMOSQRC---EGLRAIEQO-QSOGLOG--ODVEAFRTANLPSMCGV 58
 DB 102 LQCCNELHQEPLVCPTLKASAKVAKQIQOQGGKQOMVSRITYGTATHLKVCNI 161

OY 59 SPTF-CRF 65
 DB 162 PQSVYCPF 169

RESULT 11
 2SSL_PICGL
 ID 2SSL_PICGL STANDARD; PRT; 172 AA.

AC P26986; 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 2S seed storage-like protein precursor.

OC Picea glauca (White spruce).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

OX NCBI_TaxID=3330;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PG118;
 RA Newton C.H.;

RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL; X63193; CAA44875.1; -

DR PIR; S18871; S18871.

DR HSSP; P01085; IHSS.

DR InterPro; IPR003612; AAI.

DR InterPro; IPR001768; Cereal_tryp_amyl_1nh.

DR Pfam; PF00234; tryp_alpha_amyl_1.

DR SMART; SM00499; AAI; 1.

KW Seed storage protein; Albumin; Signal.

FT SIGNAL 1 35 POTENTIAL.

FT CHAIN 36 172 2S SEED STORAGE-LIKE PROTEIN.

SQ SEQUENCE 172 AA; 20393 MW; 4C29E343B61003FB CRC64;

Query Match 29.7%; Score 104; DB 1; Length 172;
 Best Local Similarity 27.1%; Pred. No. 4.2e-05;

Matches 23; Conservative 14; Mismatches 26; Indels 22; Gaps 2;

OY 1 QERSLRGCCDHLKOMOSQRCGELRAIEQOQSGOL-----QGG 40
 DB 72 RQPSERCCCEELQRMSPQCOAIOQMDQSLSYDFMDSQSDAPLNQRRRRREGGR 131

OY 41 DVFEAFRTANLPSMCGV--SPTFEC 63
 DB 132 EEEEMERAAVLPNTCNVREPPRC 156

RESULT 12
 2SSB_BRANA

ID 2SSB_BRANA STANDARD; PRT; 178 AA.

AC P27740;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Napin B precursor (1.7S seed storage protein).

GN NABP.

OS Brassica napus (Rape).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Brassica.

OX NCBI_TaxID=3708;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. SVALOF KARAT 20516-K;

RX MEDLINE=91231016; PubMed=2029903;

RA Ericson M.L., Muren E., Gustavsson H.O., Josefsson L.G., Raak L.;

RT "Analysis of the promoter region of napin genes from Brassica napus

demonstrates binding of nuclear protein in vitro to a conserved

sequence motif.";

RL Eur. J. Biochem. 197;741-746(1991).

CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPIINS ARE ONE OF THE

TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING

ITS MATURATION.

CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN

LINKED BY DISULFIDE BONDS.

CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL; X58142; CAA41150.1; -

DR PIR; S15382; S15382.

DR InterPro; IPR003612; AAI.

DR InterPro; IPR001768; Cereal_tryp_amyl_1nh.

DR InterPro; IPR000617; Napin.

DR Pfam; PF00234; tryp_alpha_amyl_1.

DR PRINTS; PR00496; NAPIIN.

DR PRODOM; PD002498; Napin; 1.

DR SMART; SM00499; AAI; 1.

KW Seed storage protein; Signal; Multigene family; Embryo.

FT SIGNAL 1 21 BY SIMILARITY.

FT PROPEP 22 38 BY SIMILARITY.

FT CHAIN 39 74 SMALL CHAIN (BY SIMILARITY).

FT PROPEP 75 94 BY SIMILARITY.

FT CHAIN 95 178 LARGE CHAIN (BY SIMILARITY).

SQ SEQUENCE 178 AA; 20114 MW; 96CE0ADB7CD966E9 CRC64;

Query Match 29.3%; Score 102.5; DB 1; Length 178;
 Best Local Similarity 38.2%; Pred. No. 6.5e-05;

Matches 26; Conservative 12; Mismatches 23; Indels 7; Gaps 4;

OY 5 LRGCCDHLKOMOSQRC---EGLRAIEQO-QSOGLOG--ODVEAFRTANLPSMCGV 58
 DB 102 LQCCNELHQEPLVCPTLKASAKVAKQIQOQGGKQKLMVSRITYGTATHLKVCNI 161

OY 59 SPTF-CRF 65
 DB 162 PQSVYCPF 169

RESULT 13
 ITRY_SINAR STANDARD; PRT; 130 AA.
 ID ITRY_SINAR
 AC P38057;
 DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin inhibitor (TISA).
OS Sinapis arvensis (Charlock).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=29728;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=94350545; PubMed=8070965;
RA Svendsen I.B., Nicolova D., Goshv I., Genov N.;
RT "Primary structure, spectroscopic and inhibitory properties of a two-
chain trypsin inhibitor from the seeds of charlock (Sinapis arvensis
L.), a member of the napin protein family."
RL Int. J. Pept. Protein Res. 43:425-430(1994).
CC -1- FUNCTION: INHIBITS TRYPSIN WITH A KI OF 7 X 10⁻⁶ M.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Multigene family; Polymorphism.
FT CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 130
FT VARIANT 32 32 LARGE CHAIN.
FT VARIANT 53 57 R -> M.
FT VARIANT 73 73 MISSING (IN FORM II).
FT VARIANT 77 77 A -> S.
FT VARIANT 81 81 K -> R.
FT VARIANT 87 81 Q -> R.
FT VARIANT 89 87 Q -> Q.
FT VARIANT 91 91 G -> H.
FT VARIANT 97 97 E -> M.
FT VARIANT 98 98 I -> V.
FT VARIANT 99 99 R -> S.
FT VARIANT 106 106 T -> K.
FT VARIANT 123 123 N -> Q.
FT VARIANT 124 124 K -> G.
FT VARIANT 126 126 M -> V.
SQ SEQUENCE 130 AA; 14682 MW; EC02E4B26D180DF2 CRC64;

Query Match 29.1%; Score 102; DB 1; Length 130;
Best Local Similarity 32.9%; Pred. No. 5.4e-05;
Matches 24; Conservative 15; Mismatches 22; Indels 12; Gaps 4;

OY 5 LRCCDHLKQMSQRC---BGLRAIEQ-QSQGLOGQ-----DVFEAFRTANLP 53
DB 50 LQCCNEHQBEPICVPTLKGAAKAVKQIQQGQGGQGLQHHEIRIRYQATHL 109
OY 54 SMCQVSPTE-CRF 65
DB 110 KVCNIPQVQCPF 122

RESULT 14
ALLI_BRAJU STANDARD; PRT; 129 AA.
AC P80207; P80215;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Allergen Bra j 1-E, small and large chains (Bra j 1).
OS Brassica juncea (leaf mustard (indian mustard)).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Sinapis arvensis (Charlock).

OC eustosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3707;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=93356721; PubMed=7688955;
RA Monsalve R.I., Gonzalez de la Pena M.A., Menendez-Arias L.,
RA Lopez-Otin C., Villalba M., Rodriguez R.;
RT "Characterization of a new oriental-mustard (Brassica juncea)
allergen, Bra j 1E: detection of an allergenic epitope."
RL Biochem. J. 293:625-632(1993)
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY TWO DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Allergen; Seed storage protein.
FT CHAIN 1 37 SMALL CHAIN.
FT NON_CONS 37 38
FT CHAIN 38 129 LARGE CHAIN.
FT VARIANT 6 6 F -> I.
FT VARIANT 20 20 R -> K.
SQ SEQUENCE 129 AA; 14644 MW; D6F2B03F62B08F8 CRC64;

Query Match 28.7%; Score 100.5; DB 1; Length 129;
Best Local Similarity 33.8%; Pred. No. 7.9e-05;
Matches 22; Conservative 15; Mismatches 17; Indels 11; Gaps 3;

OY 5 LRCCDHLKQMSQRC---BGLRAIEQ-QSQGLOGQ-----DVFEAFRTANLP 53
DB 49 LQCCNEHQBEPICVPTLKGAAKAVKQIQQGQGGQGLQHHEIRIRYQATHL 108
OY 54 SMCQV 58
DB 109 RVCNI 113

RESULT 15
2SSE_BRAJU STANDARD; PRT; 186 AA.
AC P09893;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Napin embryo specific precursor (1.7S seed storage protein).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308225; PubMed=3040733;
RA Scofield S.R., Crouch M.L.;
RT "Nucleotide sequence of a member of the napin storage protein family
from Brassica napus."
RL J. Biol. Chem. 262:12202-12208(1987).
CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC ITS MATURATION.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
CC -1- DEVELOPMENTAL STAGE: EMBRYO.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:45:37 : Search time 77.42 seconds
(without alignments)
145.242 Million cell updates/sec

Title: US-09-913-351-5
Perfect score: 350
Sequence: 1 QERSLRGCCDHLKQMSOCK.....FRITANLPSMCGVSPTECRF 65

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	347	99.1	65	10	Q9S872		Q9S872 ricinus com
2	165	47.1	139	10	P93198		P93198 juglans reg
3	158	45.1	153	10	Q9AUD1		Q9AUD1 sesamum ind
4	150.5	43.0	64	10	Q9AVK2		Q9AVK2 lycopersico
5	132.5	37.9	146	10	Q9LRC2		Q9LRC2 bertholletii
6	130.5	37.3	165	10	Q9FH31		Q9FH31 arabidopsi
7	130	37.1	141	10	Q9S649		Q9S649 cucurbita s
8	127	36.3	66	10	Q9S870		Q9S870 momordica c
9	124	35.4	148	10	Q9XHP1		Q9XHP1 sesamum ind
10	118.5	33.9	161	10	Q9L410		Q9L410 picea glauc
11	113.5	32.4	165	10	Q93YGO		Q93YGO moringa ole
12	113.5	32.4	165	10	Q40850		Q40850 picea glauc
13	112.5	32.1	139	10	Q39787		Q39787 gossypium h
14	112.5	32.1	139	10	Q39795		Q39795 gossypium h
15	110.5	31.6	323	10	Q39928		Q39928 helianthus
16	109	31.1	148	10	Q9M4E4		Q9M4E4 hordeum vul

17	108.5	31.0	148	10	Q9M4D9		Q9M4D9 tritium mo
18	107.5	30.7	167	10	Q81411		Q81411 picea glauc
19	107	30.6	148	10	Q9SC07		Q9SC07 tritium mo
20	106	30.3	147	10	Q9FS19		Q9FS19 hordeum vul
21	105	30.0	147	10	Q9AWH0		Q9AWH0 aegilops sp
22	105	30.0	148	10	Q9AMG9		Q9AMG9 aegilops sp
23	104.5	29.9	86	10	Q9S9F0		Q9S9F0 brassica na
24	104.5	29.9	178	10	Q42413		Q42413 brassica ju
25	104.5	29.9	178	10	Q39344		Q39344 brassica na
26	104	29.7	162	10	Q64929		Q64929 pseudotsuga
27	103.5	29.6	173	10	Q9ZRB1		Q9ZRB1 picea glauc
28	103	29.4	148	10	Q9AQX9		Q9AQX9 tritium ur
29	102.5	29.3	173	10	Q81412		Q81412 picea glauc
30	101.5	29.0	155	10	Q41167		Q41167 raphanus sa
31	101.5	29.0	178	10	Q42450		Q42450 brassica ol
32	101	28.9	156	10	Q941R0		Q941R0 arachis hyp
33	101	28.9	190	10	Q40995		Q40995 pinus strob
34	100	28.6	148	10	Q9M4E1		Q9M4E1 avena sativ
35	99.5	28.4	162	10	Q64932		Q64932 pseudotsuga
36	99	28.3	91	10	Q9S9E6		Q9S9E6 brassica na
37	99	28.3	91	10	Q9S9E5		Q9S9E5 brassica na
38	99	28.3	148	10	Q9AVP9		Q9AVP9 tritium mo
39	99	28.3	162	10	Q40998		Q40998 pinus strob
40	98.5	28.1	88	10	Q9S9E9		Q9S9E9 brassica na
41	98.5	28.1	164	10	Q9FV75		Q9FV75 tritium ae
42	98	28.0	152	10	Q41168		Q41168 raphanus sa
43	97.5	27.9	88	10	Q9S9E8		Q9S9E8 brassica na
44	97	27.7	172	10	Q40997		Q40997 pinus strob
45	97	27.7	174	10	Q41169		Q41169 raphanus sa

ALIGNMENTS

RESULT 1
Q9S872 PRELIMINARY: PRT: 65 AA.
ID AC Q9S872:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 7.3 KDA NAPIIN-LIKE PROTEIN LARGE CHAIN (FRAGMENT).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustoidia I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE.
RX MEDLINE=97135090; PubMed=8980648;
RA Neumann G.M., Condron R., Polya G.M.;
RT "Purification and sequencing of napin-like protein small and large
RT chains from Momordica charantia and Ricinus communis seeds and
RT determination of sites phosphorylated by plant Ca(2+)-dependent
RT protein kinase".
RL Biochim. Biophys. Acta 1298:223-240(1996).
DR InterPro: IPR003612; AAI:
DR InterPro: IPR000617; Napin.
DR PRINTS: PR00496; NAPIIN.
DR SMART: SM00499; AAI: 1.
SQ SEQUENCE 65 AA; 7350 MW; CB9264C4C917F498 CRC64;

Query Match 99.1%; Score 347; DB 10; Length 65;
Best Local Similarity 98.5%; Pred. No. 3.7e-35;
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 QERSLRGCCDHLKQMSOCKRGGLQALIEQDSQGLQGVDFEAFRRANLPSMCGVSP 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 EERSLRGCCDHLKQMSOCKRGGLQALIEQDSQGLQGVDFEAFRRANLPSMCGVSP 60
QY 61 TECRF 65
|||||

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB044391; BAA96554.1; -.
 DR HSSP: P01087; 1BIU.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1tryp_amyl_1nh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_amyl_1.
 DR PRINTS: PR00808; AMLASEINHBR.
 DR SMART: SM00496; NAPIN.
 DR SMART: SM00499; AAI; 1.
 SQ SEQUENCE 146 AA; 16910 MW; 4A69A196EE6C7096 CRC64;

Query Match 37.9%; Score 132.5; DB 10; Length 146;
 Best Local Similarity 39.7%; Pred. No. 1.3e-08;
 Matches 25; Conservative 10; Mismatches 27; Indels 1; Gaps 1;

OY 2 ESSLRGCCDHLKQMSQCRCEGLRQAIIE-QQSGGOLGQGVFEAFRTANLPSMGVSP 60
 DB 75 EPHMSCCBQLGMDSCRCCEGLRMMRMQCKEMQPREQRMRMLANLPSRCNLSP 134
 OY 61 TBC 63
 DB 135 MRC 137

RESULT 6
 O9FH31 PRELIMINARY; PRT; 165 AA.
 ID O9FH31;
 RC 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DE 01-OCT-2001 (TREMBLrel. 18, last annotation update)
 DE 2S STORAGE PROTEIN-LIKE
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:31-63(2000).
 DR EMBL: AB022214; BAB09940.1; -.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1tryp_amyl_1nh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_amyl_1.
 DR PRINTS: PR00496; NAPIN.
 DR SMART: SM00499; AAI; 1.
 SQ SEQUENCE 165 AA; 18671 MW; E93C812727F10D0F CRC64;

Query Match 37.3%; Score 130.5; DB 10; Length 165;
 Best Local Similarity 40.3%; Pred. No. 2.6e-08;
 Matches 27; Conservative 14; Mismatches 23; Indels 3; Gaps 2;

OY 1 QERSLRGCCDHLKQMSQCRCEGLRQAIIEQQSGGOLGQGVFEAFRTANLPSMGVSP 60
 DB 95 QOSSLKMCNCNEILROYDKMVCPTLKKAAQVRFQGMHGQQQVQHVFOETAKNLPVCKI-P 153
 OY 61 T--ECRF 65
 DB 154 TVGSCOF 160

RESULT 7

O39649
 ID O39649 PRELIMINARY; PRT; 141 AA.
 AC O39649;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE PREPRO25 ALBUMIN PRECURSOR.
 DE Cucurbita sp.
 OS Cucurbita sp.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Cucurbitales; Cucurbitaceae; Cucurbita.
 OC eurosids I; Cucurbitales; Cucurbitaceae; Cucurbita.
 OX NCBI_TaxID=3666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COTYLEDON;
 RX MEDLINE=92077151; PubMed=1743299;
 RA Hara-Nishimura I., Inoue K., Nishimura M.;
 RT "A unique vacuolar processing enzyme responsible for conversion of
 RT several proprotein precursors into the mature forms.";
 RL FEBS Lett. 294:89-93(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COTYLEDON;
 RX MEDLINE=94100993; PubMed=8275099;
 RA Hara-Nishimura I., Takeuchi Y., Inoue K., Nishimura M.;
 RT "Vesicle transport and processing of the precursor to 2S albumin in
 RT pumpkin.";
 RL Plant J. 4:793-800(1993).
 DR EMBL: D16560; BAA03993.1; -.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1tryp_amyl_1nh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_amyl_1.
 DR PRODOM: PD002498; Napin; 1.
 DR SMART: SM00499; AAI; 1.
 KW SIGNAL.
 FT SIGNAL.
 FT CHAIN 1 22
 FT CHAIN 36 141
 SQ SEQUENCE 141 AA; 16597 MW; 3EB12A81CE567EB5 CRC64;

Query Match 37.1%; Score 130; DB 10; Length 141;
 Best Local Similarity 39.7%; Pred. No. 2.5e-08;
 Matches 23; Conservative 11; Mismatches 22; Indels 2; Gaps 1;

OY 8 CCDHLKQMSQCRCEGLRQAIIEQQSGGOLGQGVFEAFRTANLPSMGVSPTECRF 65
 DB 86 CCEELKNVDECRCDMLFEIAREQROA--RQEGROMLQKARNLPSMGIRPORCDF 141

RESULT 8
 O9S870 PRELIMINARY; PRT; 66 AA.
 ID O9S870;
 AC O9S870;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)
 DE 7.9 KDA NAPIN-LIKE PROTEIN LARGE CHAIN (FRAGMENT).
 OS Momordica charantia (Bitter melon) (Balans pear).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Cucurbitales; Cucurbitaceae; Cucurbita.
 OC eurosids I; Cucurbitales; Cucurbitaceae; Momordica.
 OX NCBI_TaxID=3673;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=971135090; PubMed=8980648;
 RA Neumann G.M., Condron R., Polya G.M.;
 RT "Purification and sequencing of napin-like protein small and large
 RT chains from Momordica charantia and Ricinus communis seeds and
 RT determination of sites phosphorylated by plant Ca(2+)-dependent
 RT protein kinase.";
 RL Biochim. Biophys. Acta 1298:223-240(1996).
 DR InterPro: IPR003612; AAI.

DR InterPro: IPR001768; Cereal_1-tryp_aml_1nh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_aml1. 1.
 DR ProDom: PD002498; Napin. 1.
 DR SMART: SM00499; AAI. 1.
 DR SEQUENCE 66 AA; 7869 MW; 29989A217252BB90 CRC64;

Query Match 36.3%; Score 127; DB 10; Length 66;
 Best Local Similarity 35.4%; Pred. No. 2,6e-08;
 Matches 23; Conservative 15; Mismatches 25; Indels 2; Gaps 1;

QY 1 QERSLGGCCDHLKQMSOCRCGEGRLQAIQOQSGQLGQDVFEAFRTANLPSCMGVSP 60
 DB 4 ROGELECCCRQLRNYEEQCRDCLQEIAREVQROE--RQEGESQMLQKARMLPAMCGVAP 61
 QY 61 TEGRF 65
 DB 62 QRCDF 66

RESULT 9
 Q9XHP1 PRELIMINARY; PRT; 148 AA.

AC Q9XHP1;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 2S ALBUMIN.
 OS Sesamum indicum (Oriental sesame) (gingell).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Lamiales; Pedaliaceae; Sesamum.
 OC NCBI_TaxID=4182;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TAINAN 1;
 RA MEDLINE=20074970; PubMed=10606554;
 RX Tai S.S.K., Wu L.S.H., Chen E.C.F., Tzen J.T.C.;
 RT "Molecular cloning of 11S globulin and 2S albumin, the two major seed
 storage proteins in sesame."
 RL J. Agric. Food Chem. 47:4932-4938(1999).
 DR EMBL: AF091841; AAD42943.1; -.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1-tryp_aml_1nh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_aml1. 1.
 DR PRINTS: PR00496; NAPIN.
 DR SMART: SM00499; AAI. 1.
 DR SEQUENCE 148 AA; 17524 MW; BA46B033BA13E3DE CRC64;

Query Match 35.4%; Score 124; DB 10; Length 148;
 Best Local Similarity 40.0%; Pred. No. 1,4e-07;
 Matches 26; Conservative 13; Mismatches 24; Indels 2; Gaps 2;

QY 1 QERSLGGCCDHLKQMSOCRCGEGRLQAIQOQSGQLGQDVFEAFRTANLPSCMGVSP 59
 DB 79 QFEHRECCNELRDYKSHRCREALRCMRQMQQEGME-QEQMQQMQMQLPRCGMSY 137
 QY 60 PTECR 64
 DB 138 PTECR 142

RESULT 10
 O81410 PRELIMINARY; PRT; 161 AA.
 AC O81410;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE WILKMIN-LIKE STORAGE PROTEIN.

OS Picea glauca (White spruce).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
 OC NCBI_TaxID=3330;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dong J.-Z., Dunstan D.I.;
 RT "White spruce albumin-like cDNA EMB36."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF074937; AAC34613.1; -.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1-tryp_aml_1nh.
 DR Pfam: PF00234; tryp_alpha_aml1. 1.
 DR SMART: SM00499; AAI. 1.
 DR SEQUENCE 161 AA; 18595 MW; 0EE4C4DC0592FC69 CRC64;

Query Match 33.9%; Score 118.5; DB 10; Length 161;
 Best Local Similarity 33.8%; Pred. No. 7,4e-07;
 Matches 24; Conservative 11; Mismatches 21; Indels 15; Gaps 2;

QY 8 CCDHLKQMSOCRCGEGRLQAIQOQSGQLGQDVFEAFRTANLPSP 54
 DB 76 CCEELGRMSPCRCQAIQOTLEDVFMDSQDQAPLNRGQGMEEBEVLRAEELPN 135
 QY 55 MGV--SPTEC 63
 DB 136 TGNVQSPSRRC 146

RESULT 11
 Q93YG0 PRELIMINARY; PRT; 60 AA.

AC Q93YG0;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 2.1 PROTEIN (FRAGMENT).
 GN 2.1.
 OS Moringa oleifera (Horseradish tree) (Moringa pterygosperma).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eustoides II; Brassicales; Moringaceae; Moringa.
 OC NCBI_TaxID=3735;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SEED;
 RA Broin M., Kokou K., Peltier G., Joet T.;
 RT "Cloning of the cDNA coding for MO2.1 flocculating protein from
 RT Moringa oleifera and characterization of the flocculant activity of
 RT the recombinant protein."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ345072; CAC69951.1; -.
 DR NON-TER 1
 FT SEQUENCE 60 AA; 6803 MW; 9446229477B627E2 CRC64;

Query Match 32.4%; Score 113.5; DB 10; Length 60;
 Best Local Similarity 43.8%; Pred. No. 1e-06;
 Matches 21; Conservative 9; Mismatches 17; Indels 1; Gaps 1;

QY 8 CCDHLKQMSOCRCGEGRLQAIQOQSGQLGQDVFEAFRTANLPSP 54
 DB 12 CCQQLRNTSPRCRSLRQAVLTHQGGQGVGPQGVRRMYRASNIPS 59

RESULT 12
 O40850 PRELIMINARY; PRT; 165 AA.
 AC O40850;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE SEED STORAGE PROTEIN.
GN EMB25.
OS Picea glauca (White spruce).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
OX NCBI_TaxID=3330;
RN [1]
RP SEQUENCE FROM N.A.
RA Dong J.-Z., Dunstan D.I.;
RT "Gene expression during somatic embryogenesis."
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; I47745; AAB01560.1; -
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1tryp_amy1_inh.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 165 AA; 19396 MW; 69CD273982C40D20 CRC64;

Query Match 32.4%; Score 113.5; DB 10; Length 165;
Best Local Similarity 30.7%; Pred. No. 3.1e-06;
Matches 23; Conservative 14; Mismatches 19; Indels 19; Gaps 2;

QY 8 CSDHLKQMSQRCCEGLRAIDQDSQ-----GQSGQGLGQGVFEAFRTAA 50
DB 76 CCEBLQMSQRCCEGLRAIDQDSQDGPAPLNRQRGRGGRGMEEEVYRAE 135
QY 51 NLPSCGV--SPTEC 63
DB 136 ELPNTCNVRSPPRC 150

RESULT 13
Q39787 PRELIMINARY; PRT; 139 AA.
AC Q39787;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE 2S ALBUMIN STORAGE PROTEIN PRECURSOR.
GN MAT5-D.
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=3635;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COKER 201; TISSUE-COTYLEDON;
RA Galau G.A., Wang H.Y.-C., Hughes D.W.;
RT "Cotton Mat5 (Cl64) gene and cDNAs encoding a methionine-rich 2S
albumin storage protein."
RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M83301; AAA33049.1; -
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1tryp_amy1_inh.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR SMART; SM00499; AAI; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 33 59 2S ALBUMIN STORAGE PROTEIN.
FT CHAIN 64 139 2S ALBUMIN STORAGE PROTEIN.
SQ SEQUENCE 139 AA; 15831 MW; 43ACF33FE97D19B4 CRC64;

Query Match 32.1%; Score 112.5; DB 10; Length 139;
Best Local Similarity 36.1%; Pred. No. 3.4e-06;
Matches 22; Conservative 11; Mismatches 21; Indels 7; Gaps 2;

QY 8 CSDHLKQMSQRCCEGLRAIDQDSQ-----GQSGQGLGQGVFEAFRTAA 62
DB 73 CCGQLKMDTQCKGGLRATMQMOQMOGMSKQREIMQ--KYTKIMSECEMEPGR 130

QY 63 C 63
DB 131 C 131

RESULT 14
Q39795 PRELIMINARY; PRT; 139 AA.
ID Q39795
AC Q39795;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE 2S ALBUMIN STORAGE PROTEIN PRECURSOR.
GN MAT5-A.
OS Gossypium hirsutum (upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=3635;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COTYLEDON;
RA Galau G.A., Wang H.Y.-C., Hughes D.W.;
RT "Cotton Mat5-A (Cl64) gene and Mat5-D cDNAs encoding methionine-rich
2S albumin storage proteins."
RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M86213; AAA33066.1; -
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1tryp_amy1_inh.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR SMART; SM00499; AAI; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 33 59 2S ALBUMIN STORAGE PROTEIN.
FT CHAIN 64 139 2S ALBUMIN STORAGE PROTEIN.
SQ SEQUENCE 139 AA; 15700 MW; 02ACE24FFEC9EF90 CRC64;

Query Match 32.1%; Score 112.5; DB 10; Length 139;
Best Local Similarity 39.0%; Pred. No. 3.4e-06;
Matches 23; Conservative 10; Mismatches 23; Indels 3; Gaps 2;

QY 8 CSDHLKQMSQRCCEGLRAIDQDSQ-----GQSGQGLGQGVFEAFRTAA 63
DB 73 CCGQLKMDTQCKGGLRATMQMOQMOGMSKQREIMQ--KYTKIMSECEMEPGR 131

RESULT 15
Q39928 PRELIMINARY; PRT; 323 AA.
ID Q39928
AC Q39928;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE 2S ALBUMIN PRECURSOR (FRAGMENT).
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusteroideae; Asterales; Asteraceae; Asteroideae;
OX Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FLOWERING SEED;
RA Thoyts P.J.E., Millichip M., Stobart A.K., Griffiths W.T.,
RA Napier J.A., Shewry P.R.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X76101; CAA53710.1; -
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1tryp_amy1_inh.
DR Pfam; PF00234; tryp_alpha_amy1; 2.
DR PRINTS; PR00496; NAPIIN.

DR SMART; SM00499; AAI; 2.
 KW Signal.
 FT NON_TER 1 1
 FT SIGNAL <1 19
 FT CHAIN 156 323
 SQ SEQUENCE 323 AA; 38282 MW; A15378AFAD52D6B3 CRC64;

Query Match 31.6%; Score 110.5; DB 10; Length 323;
 Best Local Similarity 30.9%; Pred. No. 1.5e-05;
 Matches 21; Conservative 15; Mismatches 27; Indels 5; Gaps 1;

QY 1 QERSLRGCCDHLKQMGSGRC-----GLRQALIEDQSGOLOGDPVFEEAFRTANLPSM 55
 DB 252 EGGGLQCCCNELQNVRRRCQCEAIKEVGGQRMRRQOQRRQYGGQOTQTVERTLENLPMQ 311
 QY 56 CGVSPTEC 63
 DB 312 CDLDVQGC 319

Search completed: September 9, 2002, 12:45:37
 Job time: 294 Sec